

Université de Montréal

**Tests de permutation d'indépendance en analyse
multivariée**

par

Aurélien Guetsop Nangue

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Statistique

Novembre 2016

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

**Tests de permutation d'indépendance en analyse
multivariée**

présentée par

Aurélien Guetsop Nangue

a été évaluée par un jury composé des personnes suivantes :

François Perron

(président-rapporteur)

Martin Bilodeau

(directeur de recherche)

Christian Léger

(membre du jury)

Bharath K. Sriperumbudur

(examineur externe)

Marine Carrasco

(représentant du doyen de la FAS)

Thèse acceptée le

8 novembre 2016

DÉDICACES

À ma femme Francine Lucie et à notre fille Angela Christie.

À ma mère Victorine et à mon père Séraphin.

Table des matières

Dédicaces	iii
Liste des tableaux	vii
Table des figures	ix
Remerciements	1
Chapitre 1. Introduction	2
Bibliographie	7
Chapitre 2. Approximations to permutation tests of independence between two random vectors	10
2.1. Introduction	11
2.2. Quadratic distance between characteristic functions	13
2.2.1. The ρV coefficient	16
2.2.2. The α -distance covariance	17
2.3. Generalized coefficients	20
2.4. Quadratic distance between copulas	21
2.5. Permutation tests	22
2.5.1. Adaptive scale parameter selection of <i>HSIC</i>	24
2.5.2. Statistical packages	26
2.6. Empirical power	26
2.6.1. Empirical power of the test M_n	27
2.6.2. Empirical power of α -distance covariance tests	29
2.6.2.1. Empirical power for meta-Gaussian distributions	29
2.6.2.2. Empirical power for meta-Student distributions	32
2.6.3. Empirical power of the <i>HSIC</i> test	32
2.7. Conclusion	39

Bibliography	41
Chapter 3. Approximations to permutation tests of mutual or serial independence of random vectors	44
3.1. Introduction	46
3.2. Möbius transformation	48
3.3. Dependence statistics	48
3.3.1. Hilbert-Schmidt independence criterion	49
3.3.2. Distance covariance	50
3.3.3. Asymptotic distribution	51
3.4. A fast approximation to the permutation test	54
3.4.1. Adaptive scale parameter selection of <i>HSIC</i>	55
3.5. Combining p -values	56
3.6. Test of serial independence	57
3.7. Simulations	58
3.7.1. Dependence between three two-dimensional and two three- dimensional vectors	59
3.7.2. Empirical power for data from a Student copula	61
3.7.3. Serial dependence of a binary sequence of zeros and ones	63
3.8. Applications	65
3.8.1. Testing mutual independence between air temperature, soil temperature, humidity, wind and evaporation	65
3.8.2. Testing serial independence for financial data	67
3.9. Conclusion	68
Appendix A: Proofs	69
Appendix B: Third moment	74
Bibliography	79
Chapitre 4. Conclusion	83
Bibliographie	85
Annexe A. Programme R et C++ du premier article	A-i

Annexe B. Programme R du deuxième article.....	B-i
--	-----

Liste des tableaux

2.1	Empirical type I error rates for Gaussian distributions with the correlation matrix R_{SR} and $\rho = 0.15$	28
2.2	Empirical powers for Gaussian distributions with the correlation matrix R_{SR} and $\rho = 0.15$	28
2.3	Empirical powers of M_n for Gaussian distributions with correlation matrices R_{SR} or R_{IC} with $\rho = 0.15$ for the case $d_1 = 2$ and $d_2 = 3$. . .	29
2.4	Empirical Type I error rates of $\mathcal{V}_n^{2(\alpha)}$ using observations $Z_k^{(j)}$ for meta-Gaussian distributions with correlation matrix R_{SR} and $\rho = 0.15$	30
2.5	Powers of $\mathcal{V}_n^{2(\alpha)}$ using observations $Z_k^{(j)}$ for meta-Gaussian distributions with correlation matrix R_{SR} and $\rho = 0.15$	31
2.6	Empirical Type I error rates of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} and $\rho = 0.15$. . .	33
2.7	Powers of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} , $\rho = 0.15$, and off-diagonal block $R_{12} = 0$	34
2.8	Powers of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} and $\rho = 0.15$	35
2.9	Empirical Type I error rates of $HSIC$ for meta-Gaussian distributions. The kernel is Gaussian with scaling set at the median of distances. The null distribution of $HSIC$ is approximated using a gamma distribution as described in Gretton et al. (2008) and a Pearson type III distribution.	36
2.10	Empirical Type I error rates of $HSIC$ for meta-Student distributions with 2 degrees of freedom. The kernel is Gaussian with scaling set at the median of distances. The null distribution of $HSIC$ is approximated using a gamma distribution as described in Gretton et al. (2008) and a Pearson type III distribution.	37

2.11	Empirical type I error rates of four distance covariance tests and two <i>HSIC</i> tests for the uniform independent variables. The sample size is $n = 512$. The number of replications is 1000.....	37
2.12	Empirical type I error rates simulated from 1000 tests at 5% significance level. Four distance covariance tests and two <i>HSIC</i> tests are considered. Simulations are based on two independent Student copulas with two degrees of freedom. The sample size is $n = 100$ and the dimensions are $d_1 = d_2 = 5$	39
3.1	Empirical error rates and powers of <i>HSIC</i> and distance covariance tests of nominal level 5% between three two-dimensional and two three-dimensional vectors as in Kojadinovic and Holmes (2009). A subset is represented by the letter B . The sample size is $n = 100$. The number of replications is 1000.....	61
3.2	Empirical type I error rates simulated from 1000 tests at 5% significance level. Two distance covariance tests of index 1/2 and 1 and two <i>HSIC</i> tests of index 1/2 are considered. Simulations are based on three independent bivariate Student copulas.....	63
3.3	Empirical type I error rates for the independent sequence V_i and the empirical powers for the lag 3 dependent sequence Z_i . The tests used are the distance covariance and the adaptive <i>HSIC</i> , both of index 1. The length of the sequences is $n = 100$ each. The number of replications is 1000.....	65
3.4	Variables related to air temperature, soil temperature, humidity, wind and evaporation.....	66

Table des figures

2.1	Left panel: Power of five distance covariance tests and two <i>HSIC</i> tests for the ICA benchmark data. Right panel: Power of four distance covariance tests and two <i>HSIC</i> tests for the sinusoidally dependent data.	38
2.2	Power of four distance covariance tests and two <i>HSIC</i> tests for data sets following a Student copula with two degrees of freedom. The sample size is $n = 100$ and dimensions are $d_1 = d_2 = 5$	40
3.1	Dependogram between three two-dimensional and two three-dimensional vectors as in Kojadinovic and Holmes (2009). The sample size is $n = 100$. In the right panel, the <i>HSIC</i> test statistics are normalized by $\prod_{j \in A} \hat{\beta}_j$ as in Equation (3.8).	60
3.2	Power functions simulated from 1000 tests at 5% significance level. Two distance covariance tests of index $1/2$ and 1 are compared to two <i>HSIC</i> tests of index $1/2$. Data follow a Student copula of dimension 6 with two degrees of freedom. The sample size is $n = 100$	62
3.3	The top row exhibits dependograms for the independent sequence V_i . The bottom row contains dependograms for the sequence Z_i which is dependent at lag 3 . The tests used are the distance covariance in the left panel and the adaptive <i>HSIC</i> in the right panel, both of index 1	64
3.4	Dependograms for mutual independence of air temperature, soil temperature, relative humidity, wind and evaporation. In the right panel, the adaptive <i>HSIC</i> test statistics are normalized by $\prod_{j \in A} \hat{\beta}_j$ as in Equation (3.8).	66
3.5	Evolution of three daily financial series and their increasing rate taken at their closure value. The period of observation ranges from January 2, 2014 to March 2, 2016.	67

- 3.6 Dependograms of the serial dependence tests for the multivariate financial sequence of dimension 3. The multivariate sequence is formed by the increasing rates of the S&P/TSX composite, the S&P500 and the DOW JONES indices. The maximum lag is $p = 100$. Two tests are considered: distance covariance in the top and adaptive *HSIC* in the bottom, both of index 1. 68

REMERCIEMENTS

L'accomplissement de ce travail n'a été rendu possible que par la conjugaison de plusieurs efforts. Pendant mon cursus doctoral, j'ai bénéficié du soutien multiforme de plusieurs personnes. Trouvez en ce travail un motif de fierté.

Je tiens grandement à remercier Martin Bilodeau, mon directeur de recherche. Honnêtement, j'ai beaucoup profité de son expérience, son professionnalisme et sa disponibilité sans réserve. Son sens de l'organisation, de la précision et de la concision dans le travail ont rendu agréable mon séjour au cycle de doctorat. Je tiens aussi à le remercier pour le prestigieux soutien financier qu'il m'a apporté. Je ne saurais cacher le privilège que j'ai eu d'être encadré par lui.

J'adresse une reconnaissance sincère au personnel administratif et au corps professoral du département de mathématiques et de statistique (DMS) de l'Université de Montréal. Mes remerciements vont aussi à l'endroit du Centre de Recherches Mathématiques pour le soutien financier m'ayant permis de travailler dans un climat de sérénité et de mener à terme ma thèse. Je remercie aussi la faculté des études supérieures et postdoctorales pour la bourse accordée pendant ma dernière année de rédaction. J'adresse aussi un merci à tous mes collègues du DMS (Joseph Tagne, Joël Sango, Philippe Gagnon, Ayi Ajavon, Thierry Chekouo, Romuald Momeya, Valéry Dongmo, Romain Kadje, Fabiola Bene, Marc-Olivier Billette, Hervé Tchouaké, Bianca Côté-Brunet, etc.) pour ces années d'échanges.

Je fais un clin d'œil à mes collègues du réseau des écoles de statistique africaines, en général, et plus spécifiquement à ceux de l'Institut Sous-régional de Statistique et d'Économie Appliquée (ISSEA) de Yaoundé au Cameroun. L'espace ne me permettra pas de citer tout le monde. Je voudrais que, Zéphyrin Soh, Cyrille Batkam, Éric Lickeufack, Beaudelaire Sa'a et autres, trouvent ici l'expression de ma profonde gratitude et ma reconnaissance sans limite.

Je ne saurais terminer sans remercier ma sœur Alvine et mes frères Fabrice, Paterne, Farrel, Wilfried et Arthur, depuis le Cameroun, pour le soutien multiforme qu'ils m'apportent, malgré l'océan Atlantique qui nous sépare. Trouvez en cet accomplissement un motif d'apaisement du chagrin que vous portez de me voir si loin.

Chapitre 1

INTRODUCTION

L'analyse de la relation entre des variables aléatoires existait déjà au XIX^e siècle. Inspiré d'une idée de Sir Francis Galton dans les années 1880, Pearson (1896) proposa une méthode de calcul du coefficient de corrélation afin d'évaluer l'ampleur de la liaison linéaire entre deux variables aléatoires. Hotelling (1936) proposa d'analyser la relation entre deux ensembles de variables. Par exemple, il fit remarquer que pour mesurer la relation entre l'intelligence et l'aspect corporel d'un individu, il fallait considérer plusieurs variables pour chacun des deux concepts. L'auteur posa ainsi les jalons de l'analyse de l'association entre deux ensembles de variables. Lorsque la distribution de chaque bloc de variables aléatoires est gaussienne, la non corrélation coïncide avec l'indépendance.

Le problème de test d'hypothèses de l'indépendance entre deux vecteurs aléatoires a connu une attention considérable dans la littérature. Le chapitre 2 a pour but de recommander une procédure de test qui se comporte bien en petites et en grandes dimensions, ne nécessite pas d'hypothèses paramétriques sur la distribution des données, et s'exécute très rapidement, c'est-à-dire qu'une formule explicite pour la statistique de test existe. Certains tests statistiques comme dans Beran et al. (2007) sont calculés par résolution numérique d'un problème d'optimisation. Ces tests ne sont pas considérés dans le chapitre 2 car ils peuvent seulement être utilisés en petites dimensions et sont très fastidieux en termes de temps de calcul pour être considérés en grandes dimensions. Les statistiques qui sont considérés partagent une représentation commune à savoir qu'elles s'expriment comme le produit scalaire de matrices carrées doublement centrées d'ordre n , où n représente la taille échantillonnale. Cette riche classe de tests inclut le coefficient de corrélation RV de Escoufier (1973), la distance de covariance de Székely et al. (2007), le coefficient RV généralisé de Minas et al. (2013), la distance de covariance généralisée de Omelka et Hudecová (2013), le test de Kojadinovic et

Holmes (2009) et le critère d'indépendance de Hilbert-Schmidt (*HSIC*) de Gretton et al. (2008).

Dans la section 2.2, des tests basés sur la fonction caractéristique sont présentés. Une démonstration de la représentation de la statistique de test comme produit scalaire de matrices carrées doublement centrées est présentée. Cette démonstration est simple par rapport à d'autres preuves disponibles dans la littérature (Feuerverger, 1993; Székely et al., 2007; Gretton et al., 2008). Comme corollaire, une importante interrelation est établie entre le test de distance de covariance et le test *HSIC* avec comme noyau la fonction caractéristique d'une distribution stable. Ce noyau contient des paramètres d'échelle, et il est établi que lorsque les paramètres d'échelle convergent vers zéro, le test *HSIC* converge vers un test de distance de covariance. Ce résultat signifie que le test *HSIC* avec un noyau caractéristique d'une distribution stable d'indice α dont les paramètres d'échelle sont suffisamment petits a , pour des besoins pratiques, la même puissance qu'un test de distance de covariance d'indice α . La section 2.3 introduit des coefficients généralisés qui peuvent être définis en remplaçant des distances euclidiennes par des distances plus générales ou par des mesures de dissimilarité entre plusieurs paires échantillonales de points.

La section 2.4 introduit des tests basés sur des copules expérimentales qui sont des restrictions de fonctions de répartition sur le cube unité. Les tests sont obtenus en utilisant des rangs et incluent le test de Kojadinovic et Holmes (2009). Un autre test assez similaire et basé sur les fonctions caractéristiques est construit. Les taux d'erreur de première espèce et les puissances de ces deux tests sont comparés avec ceux du test basé sur la distance de covariance à la section 2.6. À la section 2.5, des approximations de la valeur- p pour *les tests de permutation* sont présentées. En fait, le calcul de la valeur- p exacte pour un test de permutation inclut toutes les $n!$ permutations. Certains auteurs proposent d'estimer la valeur- p par des simulations Monte Carlo où la statistique de test est recalculée pour un grand nombre de permutations. Ces tests seront référés comme des *tests randomisés*. D'autres méthodes prennent numériquement moins de temps de calcul car elles ne requièrent aucun rééchantillonnage, donc aucune permutation. Deux de ces méthodes sont considérées. Pour la classe des tests *HSIC*, Gretton et al. (2008) estiment les deux premiers moments de la distribution asymptotique du test original sous l'hypothèse nulle d'indépendance et utilisent la distribution gamma comme approximation. Kazi-Aoual et al. (1995) calculent les trois premiers moments de la distribution exacte de permutation du test basé sur le coefficient RV et utilisent une distribution de Pearson de type III pour approcher cette distribution. Minas et al. (2013) ont observé que l'approximation développée

par Kazi-Aoual et al. (1995) est en fait applicable à une famille de tests beaucoup plus générale considérée dans cette thèse. La distribution de Pearson de type III n'est rien d'autre qu'une distribution gamma relocalisée. Les outils contenant des programmes informatiques pour ces méthodes d'approximation sont présentés. Une nouvelle méthode de sélection adaptative des paramètres d'échelle du noyau caractéristique du test *HSIC* est aussi proposée. Cette méthode de sélection est un problème d'optimisation de la variance exacte du test de permutation.

La section 2.6 présente les résultats de simulations à grandes échelles. Dans un premier temps, des simulations basées sur des données gaussiennes comparent les taux d'erreur de première espèce et les puissances du test de Kojadinovic et Holmes (2009), d'un test similaire que nous avons proposé et d'un test de distance de covariance. Le test de Kojadinovic et Holmes (2009) qui est un test randomisé, a un taux d'erreur de première espèce qui approche zéro lorsque la dimension augmente et a une mauvaise puissance comparativement aux tests de distance de covariance. Dans un second temps, des simulations sont basées sur des données méta-gaussiennes et méta-student qui sont des distributions obtenues par des transformations monotones appliquées séparément à chacune des variables d'un vecteur de distribution gaussienne et student. Ces simulations montrent que les tests de distance de covariance et les tests *HSIC* ont des taux d'erreur de première espèce proches du niveau nominal de 5% en grandes dimensions. Toutes les valeurs- p sont calculées à partir de la distribution de Pearson de type III comme approximation de la distribution exacte de permutation des tests considérés. Les taux d'erreur de première espèce et les puissances des tests de α -distance de covariance sont également comparés lorsque l'indice α varie. L'approximation par la distribution de Pearson de type III conduit à des taux d'erreur de première espèce proches du niveau nominal de 5% pour presque toutes les configurations de tailles échantillonnales, de dimensions et de transformations monotones considérées, excepté quelque fois où les tailles échantillonnales sont très faibles, par exemple $n = 15$, combinées aux distributions avec de lourdes queues. L'indice α affecte les puissances et les petites valeurs de α comme $\alpha = 1/2$ ou $1/3$ conduisent généralement à des puissances proches des optimales. Dans un troisième temps, des tests de distance de covariance sont comparés aux tests *HSIC* à partir d'une simulation de référence similaire à celle de Bach et Jordan (2002) dans le contexte de l'analyse en composantes indépendantes. Il est illustré que le test *HSIC* avec un noyau caractéristique d'une distribution stable d'indice α avec un paramètre d'échelle suffisamment petit a la même puissance que le test de α -distance de covariance. Dans trois simulations dont deux sont empruntées de l'apprentissage automatique, le test *HSIC* avec la nouvelle méthode de sélection de paramètres

d'échelle présentée à la section 2.5.1 a une meilleure puissance que le test de distance de covariance.

Le chapitre 3 généralise les tests d'indépendance entre deux vecteurs aux tests d'indépendance mutuelle entre plusieurs vecteurs. Il traite aussi des tests d'indépendance sérielle d'une suite multidimensionnelle stationnaire. En fait, le problème de test d'indépendance entre p composantes d'un vecteur aléatoire a connu une attention considérable en statistique. Une approche naturelle est de considérer une fonctionnelle de la différence entre la fonction de répartition expérimentale jointe et le produit des fonctions de répartition expérimentales marginales. Cette même approche peut utiliser les fonctions caractéristiques expérimentales. Lorsque la fonctionnelle de la différence est au-dessus d'un certain seuil, les vecteurs sont déclarés dépendants. Csörgő (1985), Kankainen (1995), Sejdinovic et al. (2013b) et Fan et al. (2015) considèrent des tests d'indépendance mutuelle basés sur les fonctions caractéristiques expérimentales. Toutefois, lorsque la dépendance est détectée, il n'est pas possible d'identifier avec leurs tests, les sous-ensembles responsables de la dépendance. Cette limite est similaire à celle du test global de Fisher dans un modèle à un facteur fixe en analyse de la variance, comparé aux procédures de comparaisons multiples. Elle est aussi similaire au test d'indépendance globale du khi-deux dans un tableau de contingence à plusieurs entrées, comparé à un modèle log-linéaire avec des termes d'interaction. Pour les tests d'indépendance, une méthode utile est la transformation de Möbius.

La transformation de Möbius a une longue histoire en statistique. La transformation de Möbius d'une fonction de répartition a été proposée pour la première fois par Blum et al. (1961) pour $p = 3$. Le cas général a été traité par Deheuvels (1981), Ghoudi et al. (2001), Genest et Remillard (2005), Kojadinovic et Holmes (2009), Kojadinovic et Yan (2011) et Duchesne et al. (2012). Elle peut également être définie avec les fonctions caractéristiques comme dans Bilodeau et Lafaye de Micheaux (2005), avec les demi-espaces de probabilité comme dans Beran et al. (2007), ou avec les probabilités des cellules d'un tableau de contingence comme dans Bilodeau et Lafaye de Micheaux (2009). La première apparition de la transformation de Möbius, bien que non spécifiée explicitement, remonte aux travaux de Lancaster (1951) sur les tableaux de contingence comme cela est expliqué dans Bilodeau et Lafaye de Micheaux (2009). La communauté d'apprentissage automatique (Sejdinovic et al., 2013a) a proposé un test non paramétrique basé sur des noyaux pour l'analyse de l'interaction entre trois variables. Ce test est en fait un test basé sur la version expérimentale de la transformation de Möbius de la fonction caractéristique lorsque $p = 3$. La transformation de Möbius générale considérée dans cette thèse peut être utilisée pour construire des tests avec des

interactions plus générales et ce pour n’importe quel ordre, aussi bien pour des tests d’indépendance mutuelle que pour des tests d’indépendance sérielle.

Le chapitre 3 est organisé comme suit. La section 3.2 introduit la transformation de Möbius des fonctions caractéristiques. Elle présente une caractérisation de l’indépendance mutuelle entre p vecteurs aléatoires par la transformation de Möbius. Dans la section 3.3, de nouveaux tests basés sur la transformation de Möbius des fonctions caractéristiques expérimentales sont introduits. Ils généralisent le critère d’indépendance de Hilbert-Schmidt (Gretton et al., 2005, 2009) et le test basé sur la distance de covariance (Székely et al., 2007) au cas $p > 2$. Les nouveaux tests ont une forme commune et s’écrivent comme une somme sur tous les éléments d’une matrice exprimée comme un produit, composante par composante, de matrices doublement centrées. Une équivalence est établie entre le test *HSIC* avec des paramètres d’échelle suffisamment petits et le test de distance de covariance. La convergence faible des processus expérimentaux basés sur la transformation de Möbius est démontrée. La convergence faible du test *HSIC* est également établie. Une difficulté rencontrée dans la dérivation de la convergence faible du test de distance de covariance est décrite.

Au lieu de se baser sur la distribution limite pour conduire les tests, la section 3.4 propose l’approximation de Pearson de type III comme une approximation très rapide et précise du test de permutation. Une contribution majeure est l’obtention de l’expression exacte des trois premiers moments de la distribution de permutation. Pour les tests *HSIC* avec un noyau caractéristique, une nouvelle méthode prometteuse de sélection adaptative des paramètres d’échelle est proposée. La méthode sélectionne les paramètres d’échelle qui maximisent la variance de la distribution de permutation. Cette optimisation peut être facilement effectuée avec un algorithme de type Newton avec des dérivées numériques. Dans la section 3.5, les valeurs- p obtenues par l’approximation de Pearson de type III pour tous les sous-ensembles possibles sont combinées à la Fisher pour obtenir le test global d’indépendance mutuelle. La section 3.6 présente une adaptation de tous les résultats décrits pour la situation d’indépendance mutuelle au problème de test d’indépendance sérielle d’une suite stationnaire.

La section 3.7 adapte le dépendogramme, un graphique de Genest et Rémillard (2005), aux tests *HSIC* et de distance de covariance. Des simulations basées sur deux exemples de données continues empruntées de Kojadinovic et Holmes (2009) sont révisées. Elles illustrent la vitesse d’exécution des tests et l’amélioration de la puissance pour les tests *HSIC* avec paramètres d’échelle adaptatifs. Un autre exemple de données consiste à simuler une suite de nombres binaires avec une dépendance sérielle de délai trois. Les tests de distance de covariance et le test

HSIC détecte avec succès la dépendance introduite. Pour toutes les simulations considérées, une analyse des taux d'erreur de première espèce et des puissances des tests d'indépendance globale et pour tous les sous-ensembles possibles de vecteurs est effectuée.

Finalement, la section 3.8 contient une application aux données réelles dont les variables réfèrent à la température de l'air, à la température du sol, à l'humidité, au vent et à l'évaporation. Les tests *HSIC* ou de distance de covariance pourraient être préférés au test du ratio de vraisemblance car l'hypothèse de données gaussiennes multivariées est rejetée par le test de Henze et Zirkler (1990). Une autre application détecte une dépendance sérielle du taux de croissance journalier des séries S&P/TSX composite, S&P500 et DOW JONES, s'étalant sur la période allant du 2 janvier 2014 au 2 mars 2016. La dépendance sérielle est analysée jusqu'à un délai maximal de 99 jours. Toutes les preuves sont renvoyées aux annexes A et B du chapitre 3.

BIBLIOGRAPHIE

- F.R. Bach et M.I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3 :1–48, 2002.
- R. Beran, M. Bilodeau, et P. Lafaye de Micheaux. Nonparametric tests of independence between random vectors. *Journal of Multivariate Analysis*, 98(9) : 1805–1824, 2007.
- M. Bilodeau et P. Lafaye de Micheaux. A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, 95(2) :345–369, 2005.
- M. Bilodeau et P. Lafaye de Micheaux. A-dependence statistics for mutual and serial independence of categorical variables. *Journal of Statistical Planning and Inference*, 139(7) :2407–2419, 2009.
- J.R. Blum, J. Kiefer, et M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32(2) :485–498, 1961.
- S. Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16(3) :290–299, 1985.
- P. Deheuvels. An asymptotic decomposition for multivariate distribution-free tests of independence. *Journal of Multivariate Analysis*, 11(1) :102–113, 1981.
- P. Duchesne, K. Ghoudi, et B. Rémillard. On testing for independence between the innovations of several time series. *The Canadian Journal of Statistics*, 40 (3) :447–479, 2012.

- Y. Escoufier. Le traitement des variables vectorielles. *Biometrics*, 29(4) :751–760, 1973.
- Y. Fan, P. Lafaye de Micheaux, S. Penev, et Salopek D. Multivariate nonparametric test of independence. *Bernoulli (In revision)*, 2015.
- A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3) :419–433, 1993.
- C. Genest et B. Rémillard. Tests of independence or randomness based on the empirical copula process. *Test*, 13(2) :335–370, 2005.
- K. Ghoudi, R.J. Kulperger, et B. Rémillard. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79(2) :191–218, 2001.
- A. Gretton, O. Bousquet, A. Smola, et B. Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In S. Jain, H. Simon, et E. Tomita, editors, *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2005.
- A. Gretton, F. Kenji, C.-H. Teo, L. Song, B. Schölkopf, et A. Smola. A kernel statistical test of independence. In J.C. Platt, D. Koller, Y. Singer, et S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 585–592. Curran Associates, Inc., 2008.
- A. Gretton, K. Fukumizu, et B.K. Sriperumbudur. Discussion of : Brownian distance covariance. *The Annals of Applied Statistics*, 3(4) :1285–1294, 12 2009.
- N. Henze et B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10) :3595–3617, 1990.
- H. Hotelling. Relations between two sets of variants. *Biometrika*, 28(3/4) :321–377, 1936.
- A. Kankainen. *Consistent Testing of Total Independence Based on the Empirical Characteristic Function*. PhD thesis, University of Jyväskylä, 1995.
- F. Kazi-Aoual, S. Hitier, R. Sabatier, et J.-D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics & Data Analysis*, 20(6) :643–656, 1995.
- I. Kojadinovic et M. Holmes. Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *Journal of Multivariate Analysis*, 100(6) :1137–1154, 2009.
- I. Kojadinovic et J. Yan. Tests of serial independence for continuous multivariate time series based on a Möbius decomposition of the independence empirical copula process. *Annals of the Institute of Statistical Mathematics*, 63(2) :347–373, 2011.

- H.O. Lancaster. Complex contingency tables treated by the partition of χ^2 . *Journal of the Royal Statistical Society. Series B. Methodological*, 13(2) :242–249, 1951.
- C. Minas, E. Curry, et G. Montana. A distance-based test of association between paired heterogeneous genomic data. *Bioinformatics*, 29(20) :2555–2563, 2013.
- M. Omelka et Š. Hudecová. A comparison of the Mantel test with a generalised distance covariance test. *Environmetrics*, 24(7) :449–460, 2013.
- K. Pearson. Mathematical contributions to the theory of evolution. III. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 187 :253–318, 1896.
- D. Sejdinovic, A. Gretton, et W. Bergsma. A kernel test for three-variable interactions. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, et K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1124–1132. Curran Associates, Inc., 2013a.
- D. Sejdinovic, B. Sriperumbudur, A. Gretton, et K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5) :2263–2291, 2013b.
- G.J. Székely, M.L. Rizzo, et N.K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6) :2769–2794, 2007.

Chapitre 2

APPROXIMATIONS TO PERMUTATION TESTS OF INDEPENDENCE BETWEEN TWO RANDOM VECTORS

Cet article a été soumis à la revue *Computational Statistics and Data Analysis*.

Les principales contributions de Aurélien Guetsop Nangue à cet article sont présentées.

- Conduite d'une importante revue de la littérature.
- Conception, écriture et validation des programmes **R** et **C++**.
- Conduite des simulations.
- Rédaction d'une partie de l'article.

Approximations to permutation tests of independence between two random vectors

Martin Bilodeau and Aurélien Guetsop Nangue

Département de mathématiques et de statistique,

Université de Montréal,

C.P. 6128, Succursale A,

Montréal, Canada H3C 3J7

Abstract

The main result establishes the equivalence in terms of power between the α -distance covariance test and the Hilbert-Schmidt independence criterion (*HSIC*) test with the characteristic kernel of a stable probability distribution of index α with sufficiently small scale parameters. Large-scale simulations reveal the superiority of these two tests over other tests based on the empirical independence copula process. They also establish the usefulness of the lesser known Pearson type III approximation to the exact permutation distribution. This approximation yields tests with more accurate type I error rates than the gamma approximation usually used for *HSIC*, especially when dimensions of the two vectors are large. A new method for scale parameter selection in *HSIC* tests is proposed which improves power performance in three simulations, two of which are from machine learning.

Keywords: Characteristic function, distance covariance, Hilbert-Schmidt independence criterion, independence, permutation test, quadratic distance, *RV* coefficient.

2.1. INTRODUCTION

The problem of testing the hypothesis of independence between two random vectors has attracted considerable attention. The present study attempts to recommend a testing procedure which performs well in small samples and large dimensions, do not require parametric assumptions on the distribution, and are

easily computed. By easily computed, it is meant that an explicit formula for the test statistic exists. Some test statistics as in Beran et al. (2007) must be evaluated by numerically solving an optimization problem. These tests are not considered here since they can only be used in small dimensions and are too computationally intensive to be considered in large-scale simulations. The tests that will be considered share a common representation as an inner product between two doubly-centered square matrices of dimension n , where n is the sample size. This rich class of tests includes the RV correlation coefficient of Escoufier (1973), the distance covariance of Székely et al. (2007), the generalized RV coefficient of Minas et al. (2013), the generalized distance covariance of Omelka and Hudcová (2013), the test of Kojadinovic and Holmes (2009), and the Hilbert-Schmidt independence criterion $HSIC$ of Gretton et al. (2008).

In Section 2.2, tests based on empirical characteristic functions are presented. Other tests having the same form as the tests based on empirical characteristic functions are also presented. A proof of the inner product representation of the test statistic is presented which is simpler than other derivations found in the literature (Feuerverger, 1993; Székely et al., 2007; Gretton et al., 2008). As a corollary, an important interrelation is established between $HSIC$ with the characteristic function of a stable distribution as kernel function and distance covariance. These characteristic kernels of $HSIC$ have a scale parameter, and it is established that as the scale parameter converges to zero, the $HSIC$ test statistic converges to a distance covariance test. This holds even though distance covariance “*cannot* be used to derive a kernel-based measure of dependence” as stated in Sejdinovic et al. (2013, p. 2279). Our result means that an $HSIC$ test with the characteristic kernel of a stable distribution of index α with a sufficiently small scale parameter has, for all practical purposes, the same power as an α -distance covariance test. Section 2.3 introduces generalized coefficients which can be defined by replacing Euclidian distances by general distances or dissimilarity measures between every sample pairs of data points.

Section 2.4 introduces tests based on empirical copulas which are distribution functions on the unit hypercube. These tests are computed using ranks and include the test of Kojadinovic and Holmes (2009). Another very similar test based on characteristic functions is constructed. Type I error rates and powers of these two tests are compared with distance covariance test in Section 2.6. In Section 2.5, approximations are presented for p -values of *permutation tests* (on all $n!$ permutations). Some authors propose to estimate p -values by Monte Carlo simulations whereby the test statistic is recomputed for a large number of permutations. Such tests will be referred to as *randomization tests*. Other methods are

less numerically intensive since they do not require doing any permutation. Two such methods are considered. For *HSIC* tests, Gretton et al. (2008) estimated the first two moments of the asymptotic null distribution of the original test and used a gamma approximation. Kazi-Aoual et al. (1995) computed the exact first three moments of the permutation test and used a Pearson type III approximation which is nothing but a shifted gamma approximation. Statistical packages for these approximations are referenced. A new method for scale parameter selection in *HSIC* tests is also proposed. This adaptive method of selection is an optimization problem of the exact variance of the permutation test.

Section 2.6 presents the results of a large-scale simulation. The test of Kojadinovic and Holmes (2009) which is a randomization test, has type I error rates approaching zero as dimension increases and has very poor power compared to distance covariance tests even when distributions of data are Gaussian. On the other hand, distance covariance and *HSIC* tests have type I error rates close to the nominal level of 5% in large dimensions. Type I error rates and powers of α -distance covariance tests are also compared as the index α varies. Distributions of data considered are meta-Gaussian and meta-Student which are distributions obtained by monotone transformations of Gaussian and Student distributions. The Pearson type III approximation yields type I error rates close to 5% in almost all configurations of sample size, dimension, and monotone transformations considered, except sometimes in very small sample sizes such as $n = 15$ combined with heavy tailed distributions. The index α affects powers with small values of α such as $\alpha = 1/2$ or $1/3$ yielding powers generally close to optimal. Distance covariance is also compared to *HSIC* in a benchmark simulation similar to that of Bach and Jordan (2002) in the context of Independent Component Analysis (ICA). It is illustrated that an *HSIC* test with the characteristic kernel of a stable distribution of index α with a sufficiently small scale parameter has the same power as the α -distance covariance test. In three simulations, two of which are from machine learning, *HSIC* tests with the new adaptive selection of Section 2.5.1 has better power than distance covariance tests.

2.2. QUADRATIC DISTANCE BETWEEN CHARACTERISTIC FUNCTIONS

Consider two random vectors $Z^{(1)}$ and $Z^{(2)}$ of dimensions d_1 and d_2 , respectively. The dimension of the joint distribution of $(Z^{(1)}, Z^{(2)})$ is $d = d_1 + d_2$. Independence is characterized by characteristic functions: $Z^{(1)}$ and $Z^{(2)}$ are independent if and only if $f^{(1,2)} = f^{(1)}f^{(2)}$. Thus, a measure of dependence between

$Z^{(1)}$ and $Z^{(2)}$ is the squared modulus of the difference between the joint characteristic function $f^{(1,2)}$ and the product of marginal characteristic functions $f^{(1)}$ and $f^{(2)}$, integrated with respect to a weight function. Let $|z|$ be the modulus of the complex number z . Many such measures of dependence can be defined of the form

$$\mathcal{H}^2 = \int |f^{(1,2)}(t^{(1)}, t^{(2)}) - f^{(1)}(t^{(1)})f^{(2)}(t^{(2)})|^2 dw,$$

where the weight function

$$dw = dG^{(1)}(t^{(1)})dG^{(2)}(t^{(2)}),$$

is a product of two probability distributions. The support of the distribution $G^{(j)}$, $j = 1, 2$, is the whole Euclidian space of dimension d_j . Since characteristic functions are bounded, then the triangle inequality gives $0 \leq \mathcal{H}^2 \leq 4$ without any condition on moments. Moreover, $\mathcal{H}^2 = 0$ if and only if $Z^{(1)}$ and $Z^{(2)}$ are independent.

Let $\mathbf{Z}^{(1)}$, of dimension $n \times d_1$, and $\mathbf{Z}^{(2)}$, of dimension $n \times d_2$, be matrices of observed variables on n individuals. The corresponding test statistics use the empirical joint characteristic function

$$f_n^{(1,2)}(t^{(1)}, t^{(2)}) = \frac{1}{n} \sum_{k=1}^n \exp(i\langle t^{(1)}, Z_k^{(1)} \rangle + i\langle t^{(2)}, Z_k^{(2)} \rangle)$$

and empirical marginal characteristic functions

$$f_n^{(j)}(t^{(j)}) = \frac{1}{n} \sum_{k=1}^n \exp(i\langle t^{(j)}, Z_k^{(j)} \rangle), \quad j = 1, 2,$$

where $\langle t, s \rangle$ denote the scalar product of vectors t and s . They are given by

$$\mathcal{H}_n^2 = \int |f_n^{(1,2)}(t^{(1)}, t^{(2)}) - f_n^{(1)}(t^{(1)})f_n^{(2)}(t^{(2)})|^2 dw. \quad (2.1)$$

These statistics are easily computed using the following result which can be found in Feuerverger (1993, p. 428) in the case $d_1 = d_2 = 1$, Székely et al. (2007, p. 2776) for a certain non integrable weight function dw , or Gretton et al. (2008, p. 3) for kernel-based Hilbert-Schmidt independence criterion. Another simpler proof is given here.

Theorem 2.1. *Let $a_{kl}^{(j)} = \varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)})$, where $\varphi^{(j)}$ is the characteristic function of the distribution $G^{(j)}$. The statistic \mathcal{H}_n^2 in (2.1) can be expressed as*

$$\mathcal{H}_n^2 = \frac{1}{n^2} \text{tr} \left(A^{(1)} A^{(2)} \right) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl}^{(1)} A_{kl}^{(2)}, \quad (2.2)$$

where $A_{kl}^{(j)} = a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}$, $j = 1, 2$ and $k, l = 1, \dots, n$ and

$$\bar{a}_{k.}^{(j)} = \frac{1}{n} \sum_{l=1}^n a_{kl}^{(j)}, \quad \bar{a}_{.l}^{(j)} = \frac{1}{n} \sum_{k=1}^n a_{kl}^{(j)}, \quad \bar{a}_{..}^{(j)} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}^{(j)}.$$

PROOF. The following expression is easily verified

$$f_n^{(1,2)}(t^{(1)}, t^{(2)}) - f_n^{(1)}(t^{(1)})f_n^{(2)}(t^{(2)}) = \frac{1}{n} \sum_{k=1}^n \prod_{j=1}^2 [\exp(i\langle t^{(j)}, Z_k^{(j)} \rangle) - f_n^{(j)}(t^{(j)})]. \quad (2.3)$$

Then,

$$\begin{aligned} & \int |f_n^{(1,2)}(t^{(1)}, t^{(2)}) - f_n^{(1)}(t^{(1)})f_n^{(2)}(t^{(2)})|^2 dw \\ &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{j=1}^2 \int \left[\exp(i\langle t^{(j)}, Z_k^{(j)} - Z_l^{(j)} \rangle) - \frac{1}{n} \sum_{v=1}^n \exp(i\langle t^{(j)}, Z_k^{(j)} - Z_v^{(j)} \rangle) \right. \\ & \quad \left. - \frac{1}{n} \sum_{u=1}^n \exp(i\langle t^{(j)}, Z_u^{(j)} - Z_l^{(j)} \rangle) + \frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \exp(i\langle t^{(j)}, Z_u^{(j)} - Z_v^{(j)} \rangle) \right] dG^{(j)} t^{(j)} \\ &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{j=1}^2 [a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}], \end{aligned}$$

which proves the result. \square

Matrices $A^{(j)}$, $j = 1, 2$, in Theorem 2.1 are symmetric and doubly-centered, *i.e.* all rows and columns sum up to zero. For symmetric distributions $G^{(j)}$, they are real-valued, and if $G^{(j)}$ is invariant to orthogonal transformation, so is the test \mathcal{H}_n^2 . Many specifications for the distribution $G^{(j)}$ can be made. Let $|\cdot|_d$ be the Euclidian norm in dimension d . A choice for $G^{(j)}$ yielding simple expressions is the stable distribution of index α with characteristic function

$$\varphi^{(j)}(t^{(j)}) = \exp(-\beta_j^\alpha |t^{(j)}|_{d_j}^\alpha), \quad \alpha \in (0, 2], \quad \beta_j > 0, \quad (2.4)$$

for which

$$a_{kl}^{(j)} = \exp(-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha).$$

The test statistic \mathcal{H}_n^2 corresponding to the stable distribution of index α will be denoted $\mathcal{H}_n^{2(\alpha)}$. Gretton et al. (2008) and Gretton et al. (2005) use the theory of reproducing kernel Hilbert spaces to define a Hilbert-Schmidt independence criterion, or *HSIC*, which is given by (2.2) with doubly-centered matrices $A^{(j)}$ defined from quantities $a_{kl}^{(j)} = k^{(j)}(Z_k^{(j)}, Z_l^{(j)})$, where $k^{(j)}$ is a kernel function. It can be defined even when observations $Z_k^{(j)}$ are not necessarily vectorial. The test $\mathcal{H}_n^{2(\alpha)}$ is in fact the *HSIC* test with the characteristic kernel of the stable distribution of index α . Other statistics closely related to the form (2.2) with

doubly-centered matrices $A^{(j)}$ have been proposed. They are introduced in the following subsections.

2.2.1. The ρV coefficient

Assume the vector $(Z^{(1)}, Z^{(2)})$ has second moments and zero means. Let Σ_{11} be the covariance of $Z^{(1)}$, Σ_{22} the covariance of $Z^{(2)}$, and Σ_{12} the covariance between $Z^{(1)}$ and $Z^{(2)}$. The ρV population coefficient (Escoufier, 1973; Robert and Escoufier, 1976) is a scalar which measures the covariance (or the correlation for standardized variables) between two random vectors and it is defined as

$$\rho V = \frac{\text{tr}(\Sigma_{12}\Sigma'_{12})}{\sqrt{\text{tr}(\Sigma_{11}^2)\text{tr}(\Sigma_{22}^2)}}.$$

The hypothesis $H_0 : \rho V = 0$ is equivalent to $H_0 : \Sigma_{12} = 0$. If the two vectors are jointly Gaussian, this hypothesis tests the independence of the two vectors. Otherwise, it tests the weaker hypothesis that the two vectors are uncorrelated. In the univariate setting, *i.e.* $d_1 = d_2 = 1$, the ρV coefficient is the squared Pearson correlation. Incidentally, it also satisfies the inequality $0 \leq \rho V \leq 1$. The ρV coefficient is estimated from the sample using the RV coefficient which is now defined. When all columns of $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ have zero means, the RV coefficient is defined as

$$RV = \frac{\text{tr}(S_{12}S'_{12})}{\sqrt{\text{tr}(S_{11}^2)\text{tr}(S_{22}^2)}} = \frac{\text{tr}(\mathbf{Z}^{(1)}\mathbf{Z}^{(1)'}\mathbf{Z}^{(2)}\mathbf{Z}^{(2)'})}{\sqrt{\text{tr}[(\mathbf{Z}^{(1)}\mathbf{Z}^{(1)'})^2]\text{tr}[(\mathbf{Z}^{(2)}\mathbf{Z}^{(2)'})^2]}},$$

where $S_{ij} = (n-1)^{-1}\mathbf{Z}^{(i)}\mathbf{Z}^{(j)'}$, $i, j = 1, 2$, are sample covariance matrices. The hypothesis $H_0 : \rho V = 0$ can also be tested using only the numerator of RV , say

$$T = \text{tr}(\mathbf{Z}^{(1)}\mathbf{Z}^{(1)'}\mathbf{Z}^{(2)}\mathbf{Z}^{(2)}'), \quad (2.5)$$

as a test statistic.

When the joint distribution is Gaussian, Robert et al. (1985) derived the asymptotic distribution of RV . However, if the joint distribution has heavier tails than the Gaussian distribution, tests based on this asymptotic distribution may have a much higher level of significance than the pre-specified level of say 5%. For this reason, Cl  roux and Ducharme (1989) derived the asymptotic distribution of RV when the joint distribution is elliptical with finite fourth order moments. An example of such an elliptical distribution is the multivariate Student distribution. The hypothesis $H_0 : \rho V = 0$ is rejected at the significance level α if $nRV > c_\alpha$,

where c_α is the $1 - \alpha$ quantile of the distribution of

$$\frac{1 + \delta}{\text{tr}(\Sigma_{11}^2)\text{tr}(\Sigma_{22}^2)} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \lambda_i \mu_j Z_{ij}^2,$$

where δ is the kurtosis parameter of the elliptical distribution, λ_i is the i th eigenvalue of Σ_{11} , μ_j is the j th eigenvalue of Σ_{22} , and all Z_{ij} are independent $N(0, 1)$ variables. This test conducted with this asymptotic distribution requires large sample sizes to have a significance level close to the pre-specified level of say 5%. For Gaussian data and $d_1 = d_2 = 30$, Josse et al. (2008) have observed that the level still exceeds 5% for $n = 300$. Cl  roux et al. (1995) also proposed a test RV computed with Spearman correlations. A better alternative is to use approximate permutation tests (Kazi-Aoual et al., 1995; Josse et al., 2008) as described in Section 2.5.

2.2.2. The α -distance covariance

For $0 < \alpha < 2$, the α -distance covariance (Sz  kely et al., 2007; Sz  kely and Rizzo, 2009a) between $Z^{(1)}$ and $Z^{(2)}$ is defined as

$$\mathcal{V}^{2(\alpha)} = \int |f^{(1,2)}(t^{(1)}, t^{(2)}) - f^{(1)}(t^{(1)})f^{(2)}(t^{(2)})|^2 dw,$$

where the (non integrable) weight function is

$$dw = \frac{1}{C(d_1, \alpha)|t^{(1)}|_{d_1}^{d_1+\alpha} C(d_2, \alpha)|t^{(2)}|_{d_2}^{d_2+\alpha}} dt^{(1)} dt^{(2)}$$

and the constants are given by

$$C(d, \alpha) = \frac{2\pi^{d/2}\Gamma(1 - \alpha/2)}{\alpha 2^\alpha \Gamma((d + \alpha)/2)}.$$

It satisfies $\mathcal{V}^{2(\alpha)} \geq 0$ and $\mathcal{V}^{2(\alpha)} = 0$ if and only if $Z^{(1)}$ and $Z^{(2)}$ are independent. They also established the following inequality when $\alpha = 1$,

$$\mathcal{V}^{2(\alpha)} \leq E|Z^{(1)} - Z^{(1)*}|_{d_1}^\alpha E|Z^{(2)} - Z^{(2)*}|_{d_2}^\alpha,$$

where $Z^{(1)*}$ (resp. $Z^{(2)*}$) is an independent copy of $Z^{(1)}$ (resp. $Z^{(2)}$). The same proof establishes the general result for any α . The α -distance covariance is thus well defined for vectors with finite moments of order α , i.e. satisfying $E|Z^{(j)}|_{d_j}^\alpha < \infty$. The sample α -distance covariance is defined as

$$\mathcal{V}_n^{2(\alpha)} = \int |f_n^{(1,2)}(t^{(1)}, t^{(2)}) - f_n^{(1)}(t^{(1)})f_n^{(2)}(t^{(2)})|^2 dw.$$

They showed that the test statistic $\mathcal{V}_n^{2(\alpha)}$ can be expressed as in (2.2), but with quantities $a_{kl}^{(j)} = |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha$ not defined from characteristic functions. The proof of this representation for $\mathcal{V}_n^{2(\alpha)}$ proposed by Sz  kely et al. (2007) is long.

A more direct proof goes along the lines of the proof of Theorem 2.1 which is based on the representation (2.3) of the empirical characteristic independence process unexploited in Székely et al. (2007). Effectively, using the invariance by translation, $a_{kl}^{(j)} \mapsto a_{kl}^{(j)} - 1$, it suffices in the proof of Theorem 2.1, for the weight function defining $\mathcal{V}_n^{2(\alpha)}$, to evaluate

$$\int \frac{\exp(i\langle t^{(j)}, Z_k^{(j)} - Z_l^{(j)} \rangle) - 1}{C(d_j, \alpha) |t^{(j)}|_{d_j}^{d_j + \alpha}} dt^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha$$

with Lemma 1 of Székely et al. (2007, p. 2771).

They also defined the α -distance correlation

$$\mathcal{R}_n^{2(\alpha)} = \frac{\sum_{k,l=1}^n A_{kl}^{(1)} A_{kl}^{(2)}}{\sqrt{\sum_{k,l=1}^n (A_{kl}^{(1)})^2 \sum_{k,l=1}^n (A_{kl}^{(2)})^2}} \quad (2.6)$$

which satisfies $0 \leq \mathcal{R}_n^{2(\alpha)} \leq 1$.

The most widely used value of $\alpha = 1$ gives the standard distance covariance. In that case, $\mathcal{V}^{2(1)}$, $\mathcal{V}_n^{2(1)}$ and $\mathcal{R}_n^{2(1)}$ are denoted more simply as \mathcal{V}^2 , \mathcal{V}_n^2 and \mathcal{R}_n^2 , respectively. For testing independence, they propose the statistic $n\mathcal{V}_n^2/T_2$, where $T_2 = \bar{a}_{..}^{(1)} \bar{a}_{..}^{(2)}$. This normalization is such that the statistic has asymptotically a mean of 1. Thus, if $\mathbb{E}(|Z^{(1)}|_{d_1} + |Z^{(2)}|_{d_2}) < \infty$, then under the null hypothesis, $n\mathcal{V}_n^2/T_2$ converges in distribution to a quadratic form $Q = \sum_{j=1}^\infty \lambda_j Z_j^2$, where variables Z_j are independent $N(0, 1)$ variables, constants λ_j are non negative and depend on the distribution of $(Z^{(1)}, Z^{(2)})$, and $\mathbb{E}(Q) = 1$. The test which rejects independence for large values of $n\mathcal{V}_n^2/T_2$ is universal. It means that if $\mathbb{E}(|Z^{(1)}|_{d_1} + |Z^{(2)}|_{d_2}) < \infty$ and $\mathcal{V}^2 > 0$, then $n\mathcal{V}_n^2/T_2 \rightarrow \infty$. Cope (2009) observed that the bias of \mathcal{V}_n^2 may be substantial and increasing with dimension. Small samples of large dimension are common in genomic studies. Székely and Rizzo (2009b) came up with the unbiased estimator of \mathcal{V}^2 ,

$$U_n = \frac{n^2}{(n-1)(n-2)} \left[\mathcal{V}_n^2 - \frac{T_2}{n-1} \right], \quad n \geq 3,$$

and proposed nU_n/T_2 as a test statistic.

The following result establishes that $\mathcal{V}_n^{2(\alpha)}$ can be seen as a limiting case of $\mathcal{H}_n^{2(\alpha)}$.

Corollary 2.1. *If $G^{(j)}$ is the stable distribution of index α in the interval $(0, 2]$, $j = 1, 2$, then*

$$\lim_{\beta_1, \beta_2 \rightarrow 0} \frac{\mathcal{H}_n^{2(\alpha)}}{\beta_1^\alpha \beta_2^\alpha} = \mathcal{V}_n^{2(\alpha)}.$$

PROOF. The result follows using the invariance by translation, $a_{kl}^{(j)} \mapsto a_{kl}^{(j)} - c_j$, and the following limit,

$$\lim_{\beta_j \rightarrow 0} \frac{e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha} - 1}{\beta_j^\alpha} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha.$$

□

Corollary 2.1 shows that for sufficiently small scale parameters β_1 and β_2 , $\mathcal{H}_n^{2(\alpha)}$ (*HSIC*) becomes, up to a constant, equivalent to $\mathcal{V}_n^{2(\alpha)}$. Hence, their power functions become indistinguishable. This interrelation between *HSIC* and α -distance covariance seems to have gone unnoticed by Székely and Rizzo (2009b) and Sejdinovic et al. (2013).

Another result gives an equivalence between the distance covariance test of index $\alpha = 2$ and the *RV* coefficient. The result is the identity $\mathcal{V}_n^{2(2)} = 4T$, where T is the numerator of *RV*, which can be found in Székely et al. (2007, p. 2784) and Josse and Holmes (2014, p. 5). In fact, the numerator of the *RV* coefficient in (2.5) is rewritten as $T = \sum_{kl} Z_k^{(1)'} Z_l^{(1)} Z_k^{(2)'} Z_l^{(2)}$. Now, in $\mathcal{V}_n^{2(2)}$, there are quantities $a_{kl}^{(j)} = |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^2$ with corresponding values of $A_{kl}^{(j)} = -2Z_k^{(j)'} Z_l^{(j)}$ which proves that $\mathcal{V}_n^{2(2)} = 4T$.

The family of stable laws is not the only one leading to α -distance covariance as a limiting case. The family of multivariate Student distributions, $t_{d_j, \nu}(0, \beta_j^2 I)$, is another such family. If $G^{(j)}$ is the multivariate Student distribution, its characteristic function can be derived (by orthogonal invariance) from that of a univariate Student distribution. Hence, it can be established that the characteristic function of $G^{(j)}$ is

$$\varphi_\nu^{(j)}(t^{(j)}) = \frac{(\sqrt{\nu} \beta_j |t^{(j)}|_{d_j})^{\nu/2} K_{\nu/2}(\sqrt{\nu} \beta_j |t^{(j)}|_{d_j})}{2^{\nu/2-1} \Gamma(\nu/2)},$$

where $K_{\nu/2}$ is the Bessel function of the second type. The following limits can then be verified

$$\begin{aligned} \lim_{\beta_j \rightarrow 0} \frac{\varphi_\nu^{(j)}(t^{(j)}) - 1}{\beta_j^2} &= -\frac{\nu}{2\nu - 4} |t^{(j)}|_{d_j}^2, \quad \nu > 2, \\ \lim_{\beta_j \rightarrow 0} \frac{\varphi_\nu^{(j)}(t^{(j)}) - 1}{\beta_j^2 \log \beta_j} &= |t^{(j)}|_{d_j}^2, \quad \nu = 2, \\ \lim_{\beta_j \rightarrow 0} \frac{\varphi_\nu^{(j)}(t^{(j)}) - 1}{\beta_j^\nu} &= -c_\nu |t^{(j)}|_{d_j}^\nu, \quad 0 < \nu < 2, \end{aligned}$$

for some positive constant c_ν . Then, the limiting cases of the statistics \mathcal{H}_n^2 in (2.2) can be identified as in Corollary 2.1. In particular, for $0 < \nu = \alpha < 2$, the limiting case is the α -distance covariance, whereas, for $\nu \geq 2$, it is a multiple

of T . This all means that for small scale parameters, Student kernels with any degree of freedom $\nu > 0$ in $HSIC$ always lead to some α -distance covariance, or to T .

2.3. GENERALIZED COEFFICIENTS

Both RV and \mathcal{R}_n^2 coefficients, which are defined with Euclidian distances $|Z_k^{(j)} - Z_l^{(j)}|_{d_j}$, have been generalized to distances, also called metrics, $d^{(j)}(Z_k^{(j)}, Z_l^{(j)})$ between sample pairs $(Z_k^{(j)}, Z_l^{(j)})$, $k \neq l$, and semimetrics (a semimetric does not satisfy the triangle inequality). The purpose was to define measures of dependence when the measurements are not necessarily vectorial. One must know that zero distance covariance in general metric spaces does not always imply independence. But, this is the case for metric spaces of strong negative type (Lyons, 2013) for which a distance covariance test is universal. Sejdinovic et al. (2013) also provide a characterization of independence based on distance-induced kernels induced by semimetrics of strong negative type. For instance, the measurements might be structured as graphs or trees representing biological networks. In these applications, the semimetric is sometimes called a dissimilarity measure. The first dissimilarity measure $d^{(1)}$ could quantify genetic divergence between individuals of a tree species observed from different populations, and $d^{(2)}$ would be the geographical distance between these populations. The aim is to evaluate the relationship between geographical distance and genetic divergence (Diniz-Filho et al., 2013).

Minas et al. (2013) defined a generalized RV correlation coefficient, GRV , and Omelka and Hudecová (2013) introduced a generalized \mathcal{R}_n^2 distance correlation coefficient, \mathcal{GR}_n^2 . These two generalized coefficients of the form (2.6) with doubly-centered matrices $A^{(j)}$ defined from quantities $a_{kl}^{(j)} = d^{(j)}(Z_k^{(j)}, Z_l^{(j)})$, in fact, coincide (Omelka and Hudecová, 2013, p. 456). Unbeknownst to these authors, the machine learning community has a kernel statistical test of independence named $HSIC$ (Gretton et al., 2008, 2005), which is of the same form as a generalized coefficient \mathcal{GR}_n^2 . The Pearson type III approximation to permutation tests described in Section 2.5 will also be applicable to these generalized coefficients.

The first dissimilarity coefficient was proposed by Mantel (1967) and it is given by

$$r_M = \frac{1}{A-1} \sum_{k>l} \left(\frac{d^{(1)}(Z_k^{(1)}, Z_l^{(1)}) - \overline{d^{(1)}}}{s_{d^{(1)}}} \right) \left(\frac{d^{(2)}(Z_k^{(2)}, Z_l^{(2)}) - \overline{d^{(2)}}}{s_{d^{(2)}}} \right),$$

where $A = n(n-1)/2$,

$$\begin{aligned}\overline{d^{(j)}} &= \sum_{k>l} d^{(j)}(Z_k^{(j)}, Z_l^{(j)})/A \\ s_{d^{(1)}}^2 &= \sum_{k>l} \left(d^{(j)}(Z_k^{(j)}, Z_l^{(j)}) - \overline{d^{(j)}} \right)^2 / (A-1).\end{aligned}$$

While both r_M and \mathcal{GR}_n^2 are correlation coefficients between two dissimilarity matrices, dissimilarity measures $d^{(j)}(Z_k^{(j)}, Z_l^{(j)})$ are doubly-centered in \mathcal{GR}_n^2 , whereas they are simply centered at their overall mean in r_M . Both papers by Minas et al. (2013) and Omelka and Hudecová (2013) showed the superiority in terms of power of the generalized coefficient \mathcal{GR}_n^2 as compared to the Mantel coefficient r_M .

2.4. QUADRATIC DISTANCE BETWEEN COPULAS

The components of the vectors $(Z^{(1)}, Z^{(2)})$ are identified by defining $b_0 = 0$, $b_1 = d_1$ and $b_2 = d_1 + d_2$, so that

$$Z^{(j)} = (Z_{b_{j-1}+1}^{(j)}, \dots, Z_{b_j}^{(j)}).$$

When all variables have continuous distributions $F_s^{(j)}$ for $j = 1, 2$ and $s = b_{j-1} + 1, \dots, b_j$, the variables $U_s^{(j)} = F_s^{(j)}(Z_s^{(j)})$ are uniformly distributed over the interval $[0, 1]$. The distribution of the vector $(U^{(1)}, U^{(2)})$ is thus a copula. Since $Z^{(1)}$ and $Z^{(2)}$ are independent if and only if $U^{(1)}$ and $U^{(2)}$ are independent, independence can be characterized with a quadratic distance between copulas. Moreover, consistent estimation of copulas may be accomplished with pseudo-observations $\hat{U}_k^{(j)}$ with components $\hat{U}_{k,s}^{(j)} = R_{k,s}^{(j)}/n$, $k = 1, \dots, n$, defined with ranks (Stute, 1984). Here, for each variable indexed by s , $R_{k,s}^{(j)}$ is the rank of $Z_{k,s}^{(j)}$ among observations $Z_{1,s}^{(j)}, \dots, Z_{n,s}^{(j)}$. Tests of quadratic distance between empirical copulas given by

$$M_n = n \int_{[0,1]^d} |C_n^{(1,2)}(u^{(1)}, u^{(2)}) - C_n^{(1)}(u^{(1)})C_n^{(2)}(u^{(2)})|^2 du^{(1)} du^{(2)}$$

were proposed by Kojadinovic and Holmes (2009) and are generalizations of the tests of Genest and Rémillard (2005) when $d_1 = d_2 = 1$. They have been shown by Kojadinovic and Holmes (2009, p. 1146) to be of the form (2.2) with doubly-centered matrices $A^{(j)}$ computed from quantities

$$a_{kl}^{(j)} = \prod_{s=b_{j-1}+1}^{b_j} \left[1 - \max(\hat{U}_{k,s}^{(j)}, \hat{U}_{l,s}^{(j)}) \right]. \quad (2.7)$$

They established the consistency of the bootstrap to compute p -values. However, since ranks are used and ties can occur in bootstrap samples, they recommended to resample without replacement which amounts to doing randomization tests.

Other very similar tests can be obtained from Theorem 2.1 in which empirical characteristic functions are computed from pseudo-observations $\hat{U}_{k,s}^{(j)}$ in the interval $[0, 1]$. Consider the case $d_1 = d_2 = 1$. If f_n is the empirical characteristic function of a sample drawn from a uniform distribution on the interval $[0, 1]$ with characteristic function $f(t) = (e^{it} - 1)/(it)$, then $\text{var}[f_n(t)] = (1 - |f(t)|^2)/n$, where $|f(t)|^2 = 2[1 - \cos(t)]/t^2$. A reasonable choice for the weight function is the density $g(t) = [1 - \cos(t)]/(\pi t^2)$. More weight is then given to points t at which the variance of f_n is small. The characteristic function of the density g is $\varphi(t) = 1 - |t|$ on the interval $[-1, 1]$. This leads to quantities $a_{kl}^{(j)} = 1 - |\hat{U}_k^{(j)} - \hat{U}_l^{(j)}|$ in Theorem 2.1. This corresponds to the distance covariance test \mathcal{V}_n^2 computed from pseudo-observations. This derivation does not generalize easily to higher dimensions because very few characteristic functions of copulas can be found. The simplest derivation is obtained with the uniform distribution on the hypercube $[0, 1]^{d_j}$, which has independent components, resulting in quantities

$$a_{kl}^{(j)} = \prod_{s=b_{j-1}+1}^{b_j} [1 - |\hat{U}_{k,s}^{(j)} - \hat{U}_{l,s}^{(j)}|]. \quad (2.8)$$

Tests based on pseudo-observations are invariant to monotone transformations $Z_{k,s}^{(j)} \mapsto g_s^{(j)}(Z_{k,s}^{(j)})$. They do not assume the existence of any moment. However, they assume that all variables have continuous distributions. Distance covariance tests and characteristic kernel tests are invariant to orthogonal transformations $Z_k^{(j)} \mapsto \Gamma^{(j)} Z_k^{(j)}$ and do not assume continuous distributions. Characteristic kernel tests do not require the existence of any moment, whereas α -distance covariance tests require the existence of moments of order α . However, if $0 < \alpha < 1$, this restriction is weak since even the Cauchy distribution has a finite moment of order α .

2.5. PERMUTATION TESTS

Permutation tests of independence require only the exchangeability between observations. Suppose a permutation test of the statistic T in (2.5), or any other statistic, is wanted. The n rows of one matrix, say $\mathbf{Z}^{(2)}$, are permuted, the other matrix $\mathbf{Z}^{(1)}$ remains fixed. The permutation test does all $n!$ permutations and recomputes the statistic T for each permutation. The p -value of the permutation test is the proportion of the $n!$ recomputed values of T greater than or equal to the

observed T . The permutation test requires a considerable amount of computations for $n > 15$. Usually, a permutation test is approximated by a randomization test with a large number of permutations, say 1000. For any statistic of the form (2.2) defined with doubly-centered matrices $A^{(j)}$, the elements of the matrix $A^{(2)}$ need not be recomputed for every shuffle of $\mathbf{Z}^{(2)}$. It suffices to generate a permutation σ of the elements $1, 2, \dots, n$ and to permute accordingly the rows and columns of $A^{(2)}$ to obtain $A^{(2)}(\sigma)$ and to recompute $\text{tr}[A^{(1)}A^{(2)}(\sigma)]$.

Pearson type III and Edgeworth approximations are based on exact moments of the permutation test and can be done without actually doing any permutation. The Pearson type III method approximates the permutation distribution with a shifted gamma distribution with the same moments. The Edgeworth method uses an expansion of the cumulant generating function. In both cases, the exact moments of the permutation distribution must be computed assuming the $n!$ values of T are equiprobable.

Kazi-Aoual et al. (1995) obtained the exact first three moments of the permutation distribution of $T = \text{tr}(\mathbf{Z}^{(1)}\mathbf{Z}^{(1)'}\mathbf{Z}^{(2)}\mathbf{Z}^{(2)'})$. Minas et al. (2013, p. 4) made an important observation: expressions for moments of T derived by Kazi-Aoual et al. (1995) are applicable to any statistic of the form (2.2) defined with doubly-centered matrices $A^{(j)}$. They are thus applicable to generalized distance correlation, \mathcal{GR}_n^2 , or to Hilbert-Schmidt independence criterion, $HSIC$. Gretton et al. (2008) proposed estimates of the first two moments of the asymptotic null distribution of $HSIC$. They approximated the null distribution of $HSIC$ by the method of moments using a gamma distribution. In their text translation application with sample size $n = 50$ using $HSIC$ with a spectral kernel, this approximation did not provide an accurate model of the null distribution and they opted for the randomization test with 200 permutations. Instead, the Pearson type III approximation would incorporate a skewness correction to the null distribution of the standardized $HSIC$ and would be more accurate.

Josse et al. (2008) compared different approximations based on moments for the test T : Gaussian, log-normal, Pearson type III and Edgeworth. Regarding the skewness, they found that it is always positive, increases with the sample size, and decreases with the number of variables. The latter two approximations provided more accurate p -values. The Edgeworth approximation is more accurate for small samples. However, the Pearson type III approximation is better for large skewness, which is the case for large sample sizes and small number of variables. For the Pearson type III approximation, the permutation test of the *standardized*

variable T is approximated with a shifted gamma density given by

$$g(t) = \frac{(2/\gamma)^{4/\gamma^2}}{\Gamma(4/\gamma^2)} \left(\frac{2 + \gamma t}{\gamma} \right)^{(4-\gamma^2)/\gamma^2} e^{-2(2+\gamma t)/\gamma^2}, \quad t > -2/\gamma,$$

which has zero mean, unit variance and skewness γ . The resulting approximation is satisfying only when $\gamma > 0$. As for the Edgeworth approximation, the distribution function of the standardized statistic T is approximated by

$$G(t) = \Phi(t) - \frac{\gamma}{6}(t^2 - 1)\phi(t),$$

where ϕ and Φ are the density and distribution function of a $N(0, 1)$ distribution. A drawback of Edgeworth expansions is that the density corresponding to G may sometimes take negative values, especially for small samples.

The denominator of RV being invariant to permutations, permutation tests of RV , T , and also $\mathcal{V}_n^{2(2)}$ and $\mathcal{R}_n^{2(2)}$, are all equivalent and will report the same p -value. In this regard, the test statistic of Srivastava and Reid (2012),

$$T_1 = \frac{n \left[\text{tr}(S_{12}S'_{12}) - \frac{1}{n}\text{tr}(S_{11})\text{tr}(S_{22}) \right]}{\sqrt{\left[\text{tr}(S_{11}^2) - \frac{1}{n}\text{tr}^2(S_{11}) \right] \left[\text{tr}(S_{22}^2) - \frac{1}{n}\text{tr}^2(S_{22}) \right]}}$$

is also equivalent from the point of view of permutation tests. This same remark holds for the adjusted RV coefficient defined from the correlation matrix of Mayer et al. (2011).

Székel et al. (2007) recommended performing a randomization test of $n\mathcal{V}_n^2$ with a small number of $\lfloor 200 + 5000/n \rfloor$ permutations. From the point of view of permutation tests, by the same argument of invariance to permutations, the tests $n\mathcal{V}_n^2/T_2$, $n\mathcal{V}_n^2$, nU_n/T_2 and \mathcal{R}_n^2 are all equivalent and will yield the same p -value. Omelka and Hudecová (2013) estimated p -values of \mathcal{GR}_n^2 by randomization tests with 999 permutations. Minas et al. (2013) estimated p -values of GRV by the Pearson type III approximation. In the same way, Edgeworth and Pearson type III approximations of p -values could be computed for the test M_n of Kojadinovic and Holmes (2009). Instead, they estimated p -values from randomization tests with 1000 permutations.

2.5.1. Adaptive scale parameter selection of $HSIC$

Recall that the $HSIC$ test $\mathcal{H}_n^{2(\alpha)}$ with the characteristic kernel (2.4) of index α depends on scaling parameters β_1 and β_2 . The adaptive scale parameter selection of $HSIC$ is a major difficulty which has not been satisfactorily addressed in the literature so far. A new adaptive selection is now proposed. The statistic $n\mathcal{H}_n^{2(\alpha)}$

is distributed asymptotically under independence (Sejdinovic et al., 2013) as

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_i \mu_j Z_{i,j}^2,$$

where λ_i and μ_j are eigenvalues of certain operators, and $Z_{i,j}$ are independent standard Gaussian variables. If,

$$\max_{j \leq m} \lambda_j^2 / \sum_{j=1}^m \lambda_j^2 \rightarrow 0, \quad \max_{j \leq m} \mu_j^2 / \sum_{j=1}^m \mu_j^2 \rightarrow 0, \quad \text{as } m \rightarrow \infty,$$

then, using the Lindeberg condition, the process $n\mathcal{H}_n^{2(\alpha)}(\beta_1, \beta_2)$ indexed by (β_1, β_2) is, for large samples, approximately distributed as a bounded Gaussian process. As stated in Csörgő and Heathcote (1987): “A celebrated result of Fernique (1971) tells us roughly that the tail of the distribution of the supremum of an almost surely bounded Gaussian process behaves asymptotically as the tail of the normal distribution of the value of this process at the point which maximizes the variance function of the process.” Let $\text{var}_P[n\mathcal{H}_n^{2(\alpha)}(\beta_1, \beta_2)]$ be the exact variance of the permutation distribution available in Kazi-Aoual et al. (1995). The adaptive selection proposed is the solution of the optimization problem,

$$(\hat{\beta}_1, \hat{\beta}_2) = \arg \sup_{\beta_1 > 0, \beta_2 > 0} \text{var}_P[n\mathcal{H}_n^{2(\alpha)}(\beta_1, \beta_2)]. \quad (2.9)$$

The explicit expression of the exact permutation variance is given by the following equation

$$\begin{aligned} (n!) \text{var}_P[n\mathcal{H}_n^{2(\alpha)}(\beta_1, \beta_2)] &= (n-1)! \prod_{j=1}^2 S_2^{(j)} + (n-2)! \prod_{j=1}^2 [(T^{(j)})^2 - S_2^{(j)}] \\ &+ 2(n-2)! \prod_{j=1}^2 [T_2^{(j)} - S_2^{(j)}] + 4(n-2)! \prod_{j=1}^2 [-S_2^{(j)}] \\ &+ 4(n-3)! \prod_{j=1}^2 [-T_2^{(j)} + 2S_2^{(j)}] + 2(n-3)! \prod_{j=1}^2 [-(T^{(j)})^2 + 2S_2^{(j)}] \\ &+ (n-4)! \prod_{j=1}^2 [(T^{(j)})^2 + 2T_2^{(j)} - 6S_2^{(j)}] \\ &- (n!)^{-1} \left[(n-1)! \prod_{j=1}^2 T^{(j)} + (n-2)! \prod_{j=1}^2 [-T^{(j)}] \right]^2, \end{aligned}$$

where for $j = 1, 2$,

$$T^{(j)} = \text{tr} [A^{(j)}], \quad T_2^{(j)} = \text{tr} [(A^{(j)})^2], \quad S_2^{(j)} = \sum_k (A_{kk}^{(j)})^2,$$

and $A^{(j)}$ is doubly-centered from the matrix $a^{(j)}$ with elements $a_{kl}^{(j)} = e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha}$.

Using the above explicit expression of the exact permutation variance as a function of scale parameters, the optimization (2.9) can be rapidly achieved with a Newton-type algorithm with numerical derivatives. Performance of this adaptive selection method is investigated in Section 2.6.3 using two examples borrowed from machine learning. The test obtained with this adaptive selection is denoted $\mathcal{H}_{n,adap}^{2(\alpha)}$. Fukumizu et al. (2008) proposed to choose scale parameters such that the bootstrap variance is close to the theoretical asymptotic variance. Bootstrap variance implies resampling for each candidate of scale parameters on a grid which is computationally very costly. Moreover, this selection criterion lacks motivation.

2.5.2. Statistical packages

The randomization test of RV is implemented in the R package `ade4` (Dray and Dufour, 2007) with the function `RV.rtest`. The R package `FactoMineR` (Husson et al., 2015) has a function `coeffRV` which estimates p -values with the Pearson type III approximation. The R package `energy` (Rizzo and Szekely, 2014) computes generalized distance covariance tests with the function `dcov.test`. It approximates p -values by a randomization test. There is also the function `GRV.test` available at www2.imperial.ac.uk/~gmontana which does a generalized distance covariance test with the Pearson type III approximation. The R package `copula` (Kojadinovic and Yan, 2010) contains the function `multIndepTest` to perform the randomization test of M_n . The Matlab package `kernelIndependenceTests` which can be downloaded at people.kyb.tuebingen.mpg.de/arthur/indep.htm has two functions: `hsicTestGamma` and `hsicTestBoot`. The former uses a gamma approximation based on the first two estimated moments of the asymptotic distribution (not the permutation distribution), whereas the latter does randomization tests. This package is the work of the authors of the paper by Gretton et al. (2008).

2.6. EMPIRICAL POWER

For each case corresponding to a triplet (n, d_1, d_2) , empirical type I error rates and powers reported in all tables and figures are estimated from 1000 tests at nominal level 5%. Two structures of correlation are considered. The first structure was used by Srivastava and Reid (2012) and is given by $R_{SR} = (r_{ij})$, with $r_{ii} = 1$ and $r_{ij} = (-1)^{i+j} \rho^{|i-j|^{0.1}}$, $i \neq j$, $i, j = 1, \dots, d$. The second structure used by Kojadinovic and Holmes (2009) is the intraclass correlation denoted R_{IC}

and defined by $r_{ij} = \rho + (1 - \rho)\delta_{ij}$, where $\delta_{ij} = 1$, if $i = j$, and $\delta_{ij} = 0$, if $i \neq j$. The bounds of the rejection region of a 5% Z -test that the level is 5% based on 1000 samples would be $[0.037, 0.063]$.

This section is organized as follows. Firstly, Section 2.6.1 computes simulations based on Gaussian data and compares type I error rates and power of the test M_n of Kojadinovic and Holmes (2009), a similar test to M_n and the distance covariance test \mathcal{V}_n^2 . Secondly, the empirical type I error rates and the empirical power of the α -distance covariance tests are evaluated in Section 2.6.2. The simulations are made for meta-Gaussian distributions and meta-Student distributions with two degrees of freedom. Thirdly, empirical type I error rates are evaluated for the *HSIC* tests. The asymptotic null distribution of the *HSIC* test is approximated using a gamma distribution as described in Gretton et al. (2008) and the permutation distribution of the *HSIC* test is approximated using a Pearson type III distribution. Empirical type I error rates for the two approximations are compared. The comparisons are based on independent meta-Gaussian distributions and independent meta-Student distributions with two degrees of freedom. Again, three simulated models, two of which are from machine learning, evaluate the power performance of the new method for scale parameter selection in *HSIC* tests and the equivalence between distance covariance test and *HSIC* test with characteristic kernel of a stable distribution with a sufficiently small scale parameters.

2.6.1. Empirical power of the test M_n

The first simulation compares three tests of the form (2.2). The tests M_n of Kojadinovic and Holmes (2009) with $a_{kl}^{(j)}$ in (2.7) and the similar test \tilde{M}_n with $a_{kl}^{(j)}$ in (2.8) are based on pseudo-observations $\hat{U}_k^{(j)}$, and the distance covariance test \mathcal{V}_n^2 with $a_{kl}^{(j)} = |Z_k^{(j)} - Z_l^{(j)}|_{d_j}$ uses the original observations $Z_k^{(j)}$. The data were generated from Gaussian distributions. The value $\rho = 0.15$ was selected for both structures of correlation. For empirical type I error rates, the correlation matrix was partitioned as

$$\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

and the block R_{12} of dimension $d_1 \times d_2$ was set to zero.

Table 2.1 reports empirical type I error rates using randomization tests with 1000 permutations. It can be seen that both tests M_n and \tilde{M}_n have empirical type I error rates approaching 0 as dimensions increase and the sample size becomes greater than the maximum value of the two dimensions d_1 and d_2 . When dimension is large compared to sample size, the product structure of the quantities $a_{kl}^{(j)}$

is such that M_n and \tilde{M}_n become exactly or nearly degenerate at 0. Pearson type III approximations for M_n and \tilde{M}_n led to error messages since the exact variance of the permutation distribution became exactly or nearly 0 in higher dimensions which is the reason for resorting to Monte Carlo simulations. Empirical type I error rates of \mathcal{V}_n^2 are close to the nominal 5% level in all cases considered. In simulations reported subsequently, empirical type I error rates of \mathcal{V}_n^2 using the Pearson type III approximation are close to the nominal level in even higher dimensions.

TABLE 2.1. Empirical type I error rates for Gaussian distributions with the correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	M_n	\tilde{M}_n	\mathcal{V}_n^2	n	d_1	d_2	M_n	\tilde{M}_n	\mathcal{V}_n^2
15	1	1	0.054	0.055	0.056	50	1	1	0.053	0.053	0.051
	2	3	0.042	0.048	0.055		2	3	0.036	0.046	0.046
	5	5	0.047	0.058	0.047		5	5	0.063	0.056	0.059
	10	15	0.033	0.059	0.052		10	15	0.044	0.039	0.053
	50	50	0.001	0.005	0.050		50	50	0.043	0.033	0.042
	50	100	0	0	0.039		50	100	0.056	0	0.056
25	1	1	0.044	0.05	0.052	100	1	1	0.043	0.049	0.046
	2	3	0.058	0.053	0.047		2	3	0.068	0.053	0.048
	5	5	0.046	0.051	0.041		5	5	0.051	0.047	0.067
	10	15	0.035	0.051	0.058		10	15	0.054	0.049	0.060
	50	50	0.045	0.012	0.038		50	50	0.047	0.043	0.049
	50	100	0.012	0.001	0.046		50	100	0.055	0	0.047

TABLE 2.2. Empirical powers for Gaussian distributions with the correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	M_n	\tilde{M}_n	\mathcal{V}_n^2	n	d_1	d_2	M_n	\tilde{M}_n	\mathcal{V}_n^2
15	1	1	0.082	0.083	0.085	50	1	1	0.151	0.153	0.158
	2	3	0.058	0.077	0.097		2	3	0.038	0.228	0.275
	5	5	0.061	0.105	0.140		5	5	0.057	0.245	0.431
	10	15	0.085	0.063	0.197		10	15	0.042	0.145	0.678
25	1	1	0.096	0.099	0.106	100	1	1	0.227	0.227	0.248
	2	3	0.051	0.114	0.154		2	3	0.049	0.398	0.506
	5	5	0.065	0.131	0.231		5	5	0.05	0.508	0.764
	10	15	0.069	0.090	0.364		10	15	0.047	0.345	0.957

Empirical powers in Table 2.2 are reported only for cases (n, d_1, d_2) with small dimensions for which type I error rates were close to 5%. They show the far superiority of \mathcal{V}_n^2 over the two other tests, M_n being the least powerful. In fact, M_n has a very low power even in small dimensions $d_1 = 2$ and $d_2 = 3$ and large sample size $n = 100$. This particular case led to another small simulation to investigate the power of M_n as n increases from 100 to 1000. Powers reported in Table 2.3

show that M_n reaches a power of 0.984 by $n = 300$ when the intraclass correlation matrix R_{IC} is used. However, for R_{SR} , the power 0.356 of M_n with 1000 observations is much lower than the power 0.506 of \mathcal{V}_n^2 with only 100 observations in Table 2.2. Considering the poor performance of the tests M_n and \tilde{M}_n , except for $d_1 = d_2 = 1$, these two tests were excluded from subsequent simulations.

TABLE 2.3. Empirical powers of M_n for Gaussian distributions with correlation matrices R_{SR} or R_{IC} with $\rho = 0.15$ for the case $d_1 = 2$ and $d_2 = 3$.

	n					
	100	200	300	400	500	1000
R_{SR}	0.072	0.082	0.116	0.133	0.153	0.356
R_{IC}	0.691	0.925	0.984	0.999	1	1

All tests in the following sections are computed from the original observations $Z_k^{(j)}$ and not on the pseudo-observations (ranks) $\hat{U}_k^{(j)}$.

2.6.2. Empirical power of α -distance covariance tests

Empirical powers of $\mathcal{V}_n^{2(\alpha)}$ were simulated for indices $\alpha = 1/2, 1, 3/2, 2$. It should be recalled that for permutation tests, $\mathcal{V}_n^{2(2)}$ is equivalent to RV . The joint distribution of $(Z_k^{(1)}, Z_k^{(2)})$ was meta-elliptical, with either a Gaussian or Student copula. The Pearson type III approximation was used for computing all p -values. The correlation structure used was R_{SR} , leading to a dependence structure more difficult to detect than that of R_{IC} . The relatively small value of $\rho = 0.15$ was used so that powers do not approach one too rapidly as n , d_1 , and d_2 increase. The (n, d_1, d_2) cases considered were the same as in Srivastava and Reid (2012) and cover a large spectrum of sample sizes and dimensions. The same monotone transformation, among three transformations, was applied to all variables resulting in three meta-distributions:

- (1) The identity transformation, $z \mapsto z$.
- (2) The square root transformation, $z \mapsto \text{sgn}(z)|z|^{1/2}$.
- (3) The cubic square root transformation, $z \mapsto \text{sgn}(z)|z|^{3/2}$.

2.6.2.1. Empirical power for meta-Gaussian distributions

Data were generated from Gaussian distributions, $N_d(0, R_{SR})$, and transformed to obtain empirical powers. All variables have the same distribution with excess kurtoses $\kappa = 0$ for the identity transformation, $\kappa = \pi/2 - 3 \approx -1.43$ (sub-Gaussian or platykurtic) for the square root transformation, and $\kappa = 15\pi/8 - 3 \approx$

2.89 (super-Gaussian or leptokurtic) for the cubic square root transformation. For empirical type I error rates, the off-diagonal block R_{12} of R_{SR} was set to 0. The Gaussian likelihood ratio test was not considered since its power is low compared to RV (Srivastava and Reid, 2012), even for Gaussian distributions. Moreover, it is applicable only when $n > d$.

TABLE 2.4. Empirical Type I error rates of $\mathcal{V}_n^{2(\alpha)}$ using observations $Z_k^{(j)}$ for meta-Gaussian distributions with correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	Transformations											
			Identity				$\text{sgn}(z) z ^{1/2}$				$\text{sgn}(z) z ^{3/2}$			
			α				α				α			
			$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
15	2	3	0.050	0.056	0.061	0.056	0.058	0.053	0.052	0.053	0.045	0.047	0.054	0.058
	5	5	0.051	0.053	0.058	0.052	0.041	0.044	0.048	0.048	0.050	0.058	0.054	0.053
	10	15	0.042	0.040	0.038	0.040	0.036	0.037	0.040	0.043	0.035	0.039	0.045	0.045
	50	50	0.051	0.052	0.051	0.052	0.042	0.043	0.043	0.043	0.054	0.052	0.053	0.049
	50	100	0.048	0.045	0.045	0.043	0.048	0.049	0.050	0.050	0.045	0.044	0.045	0.042
	100	200	0.041	0.040	0.039	0.040	0.041	0.041	0.041	0.043	0.043	0.042	0.042	0.040
	200	300	0.036	0.035	0.036	0.033	0.050	0.049	0.048	0.046	0.041	0.039	0.036	0.037
	400	600	0.053	0.051	0.050	0.049	0.056	0.057	0.059	0.057	0.046	0.047	0.048	0.047
25	2	3	0.051	0.054	0.050	0.049	0.050	0.055	0.061	0.061	0.046	0.043	0.048	0.050
	5	5	0.047	0.048	0.046	0.049	0.050	0.044	0.039	0.044	0.046	0.041	0.051	0.049
	10	15	0.056	0.057	0.063	0.065	0.061	0.055	0.055	0.054	0.054	0.060	0.063	0.063
	50	50	0.041	0.045	0.044	0.044	0.042	0.041	0.041	0.042	0.053	0.053	0.055	0.052
	50	100	0.038	0.039	0.037	0.036	0.046	0.045	0.044	0.044	0.041	0.042	0.045	0.043
	100	200	0.058	0.054	0.052	0.051	0.060	0.058	0.058	0.058	0.049	0.049	0.047	0.045
	200	300	0.057	0.058	0.059	0.060	0.053	0.053	0.054	0.057	0.049	0.051	0.053	0.052
	400	600	0.050	0.052	0.052	0.052	0.056	0.054	0.058	0.058	0.049	0.049	0.050	0.049
50	2	3	0.050	0.047	0.039	0.041	0.042	0.042	0.044	0.037	0.050	0.043	0.044	0.048
	5	5	0.050	0.041	0.042	0.046	0.052	0.053	0.049	0.051	0.039	0.042	0.038	0.029
	10	15	0.051	0.053	0.050	0.046	0.054	0.054	0.055	0.053	0.039	0.037	0.042	0.042
	50	50	0.056	0.052	0.050	0.050	0.055	0.054	0.056	0.056	0.058	0.053	0.048	0.044
	50	100	0.056	0.056	0.056	0.059	0.051	0.054	0.056	0.056	0.060	0.062	0.062	0.060
	100	200	0.047	0.048	0.046	0.047	0.053	0.053	0.056	0.057	0.042	0.042	0.044	0.042
	200	300	0.049	0.046	0.048	0.049	0.047	0.047	0.049	0.050	0.058	0.059	0.061	0.057
	400	600	0.042	0.043	0.044	0.045	0.051	0.049	0.048	0.049	0.045	0.044	0.043	0.040
100	2	3	0.053	0.051	0.048	0.046	0.057	0.056	0.054	0.051	0.053	0.048	0.044	0.039
	5	5	0.049	0.057	0.058	0.055	0.055	0.050	0.050	0.053	0.053	0.055	0.051	0.049
	10	15	0.042	0.038	0.035	0.029	0.040	0.035	0.038	0.035	0.035	0.036	0.037	0.042
	50	50	0.039	0.038	0.041	0.042	0.049	0.051	0.049	0.050	0.041	0.039	0.033	0.033
	50	100	0.074	0.074	0.073	0.077	0.063	0.062	0.060	0.058	0.070	0.075	0.073	0.073
	100	200	0.043	0.043	0.039	0.038	0.049	0.048	0.049	0.051	0.050	0.050	0.041	0.040
	200	300	0.045	0.045	0.043	0.044	0.048	0.049	0.048	0.049	0.045	0.044	0.043	0.040
	400	600	0.051	0.050	0.049	0.046	0.055	0.055	0.055	0.054	0.050	0.050	0.050	0.047

TABLE 2.5. Powers of $\mathcal{V}_n^{2(\alpha)}$ using observations $Z_k^{(j)}$ for meta-Gaussian distributions with correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	Transformations											
			Identity				$\text{sgn}(z) z ^{1/2}$				$\text{sgn}(z) z ^{3/2}$			
			α				α				α			
			$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
15	2	3	0.089	0.101	0.107	0.114	0.076	0.081	0.094	0.100	0.095	0.107	0.112	0.114
	5	5	0.120	0.125	0.133	0.144	0.111	0.117	0.132	0.128	0.103	0.112	0.127	0.127
	10	15	0.169	0.181	0.197	0.206	0.150	0.164	0.173	0.175	0.145	0.162	0.171	0.178
	50	50	0.325	0.336	0.349	0.364	0.317	0.319	0.320	0.324	0.272	0.288	0.306	0.316
	50	100	0.316	0.326	0.341	0.354	0.287	0.298	0.305	0.311	0.291	0.308	0.324	0.341
	100	200	0.404	0.411	0.422	0.430	0.392	0.397	0.399	0.401	0.354	0.363	0.372	0.386
	200	300	0.480	0.494	0.502	0.507	0.454	0.461	0.463	0.468	0.421	0.433	0.443	0.453
	400	600	0.471	0.477	0.486	0.487	0.442	0.442	0.442	0.444	0.428	0.437	0.442	0.446
25	2	3	0.131	0.156	0.167	0.179	0.133	0.144	0.155	0.166	0.111	0.136	0.144	0.148
	5	5	0.206	0.226	0.242	0.257	0.187	0.211	0.218	0.222	0.180	0.203	0.220	0.230
	10	15	0.303	0.323	0.349	0.371	0.281	0.298	0.310	0.318	0.246	0.279	0.308	0.323
	50	50	0.557	0.577	0.588	0.598	0.530	0.538	0.553	0.566	0.491	0.520	0.537	0.548
	50	100	0.611	0.632	0.642	0.648	0.555	0.574	0.587	0.604	0.539	0.560	0.574	0.587
	100	200	0.675	0.686	0.695	0.711	0.653	0.660	0.664	0.669	0.615	0.634	0.645	0.655
	200	300	0.714	0.721	0.728	0.731	0.685	0.690	0.695	0.699	0.645	0.660	0.672	0.680
	400	600	0.752	0.757	0.762	0.771	0.735	0.738	0.741	0.746	0.721	0.735	0.742	0.746
50	2	3	0.208	0.247	0.289	0.297	0.193	0.220	0.239	0.256	0.174	0.238	0.260	0.252
	5	5	0.383	0.442	0.483	0.499	0.351	0.394	0.413	0.430	0.335	0.410	0.445	0.442
	10	15	0.651	0.691	0.722	0.745	0.606	0.636	0.656	0.678	0.560	0.623	0.650	0.661
	50	50	0.884	0.897	0.909	0.915	0.868	0.876	0.884	0.894	0.846	0.863	0.878	0.884
	50	100	0.925	0.937	0.945	0.948	0.910	0.915	0.921	0.924	0.888	0.901	0.915	0.921
	100	200	0.950	0.954	0.957	0.962	0.945	0.946	0.947	0.951	0.934	0.938	0.943	0.950
	200	300	0.969	0.971	0.975	0.978	0.963	0.964	0.968	0.969	0.949	0.956	0.958	0.961
	400	600	0.975	0.976	0.977	0.979	0.970	0.972	0.974	0.975	0.970	0.974	0.974	0.975
100	2	3	0.426	0.497	0.537	0.556	0.369	0.431	0.456	0.491	0.369	0.458	0.513	0.507
	5	5	0.731	0.782	0.808	0.826	0.692	0.732	0.757	0.784	0.646	0.730	0.771	0.778
	10	15	0.930	0.954	0.965	0.969	0.909	0.925	0.934	0.939	0.889	0.930	0.948	0.955
	50	50	0.997	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.994	0.995	0.996	0.997
	50	100	0.999	1	1	1	1	1	1	1	0.998	0.998	0.999	0.999
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	0.999	0.999	0.999	0.999
	400	600	1	1	1	1	1	1	1	1	1	1	1	1

Table 2.4 shows that tests $\mathcal{V}_n^{2(\alpha)}$ have empirical type I error rates close to 5% for all indices α and cases (n, d_1, d_2) considered. The Pearson type III approximation was very successful at estimating the distribution of the permutation test. It is also faster than doing a randomization test with 1000 permutations. In Table 2.5, although $\mathcal{V}_n^{2(2)}$ is most powerful and the power decreases with α , tests using smaller values of α do not yield much lower power. Thus, any of the α -distance covariance tests yields a reasonable testing procedure for the meta-Gaussian distributions considered here.

2.6.2.2. Empirical power for meta-Student distributions

Data were generated from Student distributions with two degrees of freedom, $t_{d,2}(0, R_{SR})$, and transformed to obtain empirical powers. All variables have the same distribution with undefined excess kurtoses since moments of order four do not exist. However the square root transformation has moments of order $\alpha < 4$, the identity transformation has moments of order $\alpha < 2$, and the cubic square root transformation has moments of order $\alpha < 4/3$. For empirical powers, data were generated not only for R_{SR} with $\rho = 0.15$, but also for R_{SR} having an off-diagonal block $R_{12} = 0$, since it is also a situation of dependence. For empirical type I error rates, it was not sufficient to set $R_{12} = 0$. Instead, data were simulated independently from Student distributions $t_{d_1,2}(0, R_{SR})$ and $t_{d_2,2}(0, R_{SR})$.

Empirical type I error rates for three transformations in Table 2.6 are not all close to 5%. For the smallest sample size $n = 15$, these rates are generally higher than 5%, especially for the cubic square root transformation where rates can be as high as 9.4%. For $n \geq 25$, the rates are generally close to 5%.

Empirical powers in Tables 2.7 and 2.8 show that the smaller value $\alpha = 1/2$ yields higher powers for the identity and cubic square root transformations which have heavy tails. For the square root transformation which has light tails, an intermediate value of $\alpha = 1$ yields powers not far from the optimal one. In this case, the value $\alpha = 1/2$ produces smaller powers but only when sample sizes and dimensions are small, otherwise, powers do not differ significantly.

2.6.3. Empirical power of the *HSIC* test

Empirical type I error rates of *HSIC* with the Gaussian kernel, with scale parameter

$$\beta_j = 1/\text{med}_{k < l} |Z_k^{(j)} - Z_l^{(j)}| \quad (2.10)$$

suggested by Gretton et al. (2008), were simulated using the function `hsicTestGamma` of the Matlab package `kernelIndependenceTests`. The null distribution in `hsicTestGamma` is obtained by matching the first two estimated moments of the asymptotic distribution of *HSIC* to those of a gamma distribution. Table 2.9 reports error rates when two vectors follow independent meta-Gaussian distributions obtained by transforming the variables of $Z^{(1)} \sim N_{d_1}(0, I)$ and $Z^{(2)} \sim N_{d_2}(0, I)$. Table 2.10 contains similar error rates when two vectors follow independent meta-Student distributions (two degrees of freedom) obtained by transforming the variables of $Z^{(1)} \sim t_{d_1,2}(0, I)$ and $Z^{(2)} \sim t_{d_2,2}(0, I)$. These error rates are far from 5% even when dimensions are moderate. Error rates in Gretton et al. (2008) are only for cases where the sample size is very large

TABLE 2.6. Empirical Type I error rates of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	Transformations											
			Identity				$\text{sgn}(z) z ^{1/2}$				$\text{sgn}(z) z ^{3/2}$			
			α				α				α			
			$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
15	2	3	0.062	0.065	0.069	0.069	0.046	0.052	0.052	0.045	0.073	0.085	0.085	0.085
	5	5	0.058	0.055	0.061	0.062	0.049	0.051	0.055	0.056	0.063	0.071	0.087	0.084
	10	15	0.068	0.071	0.078	0.082	0.057	0.062	0.065	0.069	0.072	0.077	0.079	0.084
	50	50	0.074	0.077	0.081	0.081	0.063	0.066	0.073	0.073	0.077	0.079	0.081	0.081
	50	100	0.069	0.073	0.075	0.078	0.059	0.059	0.063	0.065	0.072	0.077	0.080	0.081
	100	200	0.080	0.082	0.091	0.086	0.074	0.077	0.076	0.073	0.083	0.094	0.090	0.088
	200	300	0.058	0.065	0.068	0.068	0.049	0.053	0.057	0.065	0.062	0.069	0.071	0.075
	400	600	0.063	0.067	0.063	0.063	0.058	0.057	0.062	0.065	0.065	0.065	0.063	0.066
25	2	3	0.047	0.055	0.061	0.064	0.048	0.046	0.045	0.047	0.053	0.057	0.061	0.059
	5	5	0.052	0.052	0.054	0.060	0.052	0.048	0.048	0.054	0.052	0.057	0.056	0.053
	10	15	0.063	0.069	0.070	0.063	0.042	0.043	0.052	0.060	0.065	0.069	0.066	0.061
	50	50	0.050	0.051	0.049	0.046	0.046	0.047	0.047	0.047	0.050	0.048	0.049	0.049
	50	100	0.069	0.066	0.066	0.067	0.063	0.063	0.059	0.061	0.066	0.067	0.064	0.063
	100	200	0.047	0.044	0.044	0.044	0.049	0.051	0.046	0.046	0.046	0.046	0.043	0.043
	200	300	0.066	0.076	0.073	0.071	0.069	0.070	0.068	0.067	0.067	0.076	0.069	0.070
	400	600	0.064	0.064	0.066	0.064	0.055	0.055	0.061	0.064	0.061	0.067	0.063	0.060
50	2	3	0.052	0.043	0.048	0.040	0.054	0.059	0.056	0.052	0.049	0.041	0.038	0.038
	5	5	0.055	0.055	0.042	0.044	0.048	0.052	0.046	0.047	0.058	0.047	0.045	0.047
	10	15	0.048	0.050	0.048	0.048	0.045	0.048	0.044	0.043	0.050	0.051	0.048	0.048
	50	50	0.051	0.049	0.054	0.053	0.041	0.044	0.047	0.049	0.056	0.054	0.047	0.047
	50	100	0.058	0.064	0.065	0.064	0.056	0.056	0.057	0.058	0.062	0.062	0.058	0.057
	100	200	0.065	0.064	0.063	0.061	0.058	0.063	0.065	0.065	0.064	0.064	0.060	0.059
	200	300	0.051	0.054	0.056	0.055	0.044	0.047	0.049	0.054	0.053	0.055	0.053	0.053
	400	600	0.063	0.063	0.058	0.061	0.061	0.060	0.062	0.061	0.062	0.060	0.062	0.060
100	2	3	0.044	0.044	0.040	0.035	0.053	0.051	0.048	0.040	0.046	0.050	0.045	0.040
	5	5	0.070	0.059	0.057	0.057	0.053	0.049	0.050	0.045	0.060	0.055	0.051	0.049
	10	15	0.051	0.048	0.044	0.040	0.049	0.048	0.051	0.049	0.048	0.042	0.040	0.034
	50	50	0.059	0.054	0.047	0.043	0.056	0.057	0.055	0.051	0.051	0.045	0.040	0.032
	50	100	0.045	0.044	0.042	0.035	0.049	0.043	0.038	0.039	0.040	0.042	0.034	0.028
	100	200	0.045	0.050	0.049	0.047	0.042	0.044	0.048	0.049	0.047	0.050	0.046	0.039
	200	300	0.058	0.051	0.054	0.050	0.059	0.062	0.054	0.051	0.057	0.053	0.048	0.042
	400	600	0.042	0.044	0.039	0.040	0.038	0.040	0.044	0.044	0.043	0.039	0.037	0.032

compared to dimensions. In fact, the largest dimension considered was four with sample sizes 1024 or higher. Error rates for the Pearson type III (shifted gamma) approximation are all close to 5% and provide a far more accurate approximation than the gamma approximation of Gretton et al. (2008).

Empirical powers were simulated as in Gretton et al. (2008), Gretton et al. (2009) and Sejdinovic et al. (2013). All tests were conducted using the Pearson type III approximation which is more reliable than the gamma approximation of

TABLE 2.7. Powers of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} , $\rho = 0.15$, and off-diagonal block $R_{12} = 0$.

n	d_1	d_2	Transformations											
			Identity				$\text{sgn}(z) z ^{1/2}$				$\text{sgn}(z) z ^{3/2}$			
			α				α				α			
			$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
15	2	3	0.618	0.583	0.523	0.483	0.166	0.199	0.229	0.245	0.706	0.627	0.566	0.531
	5	5	0.884	0.850	0.790	0.726	0.388	0.415	0.416	0.413	0.936	0.888	0.809	0.754
	10	15	0.985	0.977	0.962	0.927	0.781	0.773	0.760	0.733	0.993	0.980	0.960	0.917
	50	50	1	1	1	0.999	0.978	0.975	0.969	0.960	1	1	0.999	0.997
	50	100	1	1	1	1	0.998	0.997	0.996	0.994	1	1	1	1
	100	200	1	1	1	1	0.999	0.998	0.998	0.998	1	1	1	1
	200	300	1	1	1	1	0.999	0.999	0.999	0.999	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
25	2	3	0.774	0.728	0.646	0.572	0.198	0.235	0.266	0.268	0.871	0.779	0.683	0.609
	5	5	0.965	0.949	0.915	0.834	0.474	0.514	0.504	0.489	0.981	0.959	0.899	0.834
	10	15	1	1	0.997	0.987	0.853	0.847	0.833	0.798	1	1	0.998	0.980
	50	50	1	1	1	1	0.999	0.999	0.996	0.995	1	1	1	1
	50	100	1	1	1	1	0.999	0.999	0.999	0.999	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
50	2	3	0.948	0.928	0.832	0.707	0.266	0.305	0.327	0.341	0.989	0.949	0.830	0.725
	5	5	0.997	0.996	0.983	0.928	0.652	0.640	0.612	0.591	1	0.997	0.980	0.932
	10	15	1	1	1	0.999	0.962	0.951	0.937	0.921	1	1	1	0.999
	50	50	1	1	1	1	1	1	1	1	1	1	1	1
	50	100	1	1	1	1	1	1	1	1	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
100	2	3	1	0.999	0.966	0.850	0.404	0.414	0.402	0.410	1	0.997	0.962	0.875
	5	5	1	1	0.999	0.980	0.863	0.804	0.742	0.694	1	1	0.995	0.978
	10	15	1	1	1	1	1	0.998	0.991	0.979	1	1	1	1
	50	50	1	1	1	1	1	1	1	1	1	1	1	1
	50	100	1	1	1	1	1	1	1	1	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1

Gretton et al. (2008). The following description is an excerpt from Sejdinovic et al. (2013): “we generated univariate random variables from the Independent Component Analysis (ICA) benchmark densities of Bach and Jordan (2002); rotated them in the product space by an angle θ between 0 and $\pi/4$ to introduce dependence; filled additional dimensions with independent Gaussian noise; and, finally, passed the resulting multivariate data through random and independent orthogonal transformations.” The two resulting random vectors were dependent

TABLE 2.8. Powers of $\mathcal{V}_n^{2(\alpha)}$ for meta-student distributions with two degrees of freedom, correlation matrix R_{SR} and $\rho = 0.15$.

n	d_1	d_2	Identity				Transformations $\text{sgn}(z) z ^{1/2}$				$\text{sgn}(z) z ^{3/2}$			
			α				α				α			
			$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
15	2	3	0.665	0.636	0.572	0.526	0.242	0.270	0.294	0.303	0.764	0.680	0.600	0.552
	5	5	0.902	0.867	0.821	0.748	0.485	0.516	0.530	0.512	0.936	0.877	0.796	0.742
	10	15	0.993	0.988	0.976	0.952	0.835	0.837	0.829	0.801	0.993	0.989	0.969	0.937
	50	50	1	1	1	1	0.996	0.996	0.994	0.993	1	1	1	1
	50	100	1	1	1	1	1	1	1	0.998	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
25	2	3	0.845	0.818	0.710	0.621	0.303	0.361	0.397	0.397	0.911	0.821	0.728	0.649
	5	5	0.985	0.973	0.939	0.875	0.652	0.669	0.655	0.635	0.994	0.978	0.928	0.852
	10	15	0.999	0.998	0.996	0.988	0.946	0.943	0.934	0.918	1	0.998	0.989	0.971
	50	50	1	1	1	1	0.999	0.999	0.999	0.999	1	1	1	1
	50	100	1	1	1	1	1	1	1	1	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
50	2	3	0.973	0.965	0.901	0.811	0.507	0.559	0.575	0.561	0.996	0.967	0.875	0.777
	5	5	1	1	0.993	0.961	0.854	0.865	0.850	0.805	1	0.999	0.980	0.938
	10	15	1	1	1	1	1	0.998	0.994	0.988	1	1	1	1
	50	50	1	1	1	1	1	1	1	1	1	1	1	1
	50	100	1	1	1	1	1	1	1	1	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1
100	2	3	0.999	0.998	0.983	0.902	0.791	0.792	0.783	0.728	1	0.999	0.973	0.892
	5	5	1	1	1	0.997	0.992	0.987	0.973	0.951	1	1	1	0.989
	10	15	1	1	1	1	1	1	1	1	1	1	1	1
	50	50	1	1	1	1	1	1	1	1	1	1	1	1
	50	100	1	1	1	1	1	1	1	1	1	1	1	1
	100	200	1	1	1	1	1	1	1	1	1	1	1	1
	200	300	1	1	1	1	1	1	1	1	1	1	1	1
	400	600	1	1	1	1	1	1	1	1	1	1	1	1

but uncorrelated, except for the case $\theta = 0$, which corresponds to independence between the two random vectors. The simulation reported here has only 5 benchmark densities as opposed to 18 in the ICA benchmark, but will serve to illustrate our point. The 5 densities, with their excess kurtoses, are Gaussian ($\kappa = 0$), Student with 3 degrees of freedom (κ undefined), Student with 5 degrees of freedom ($\kappa = 6$), uniform ($\kappa = -1.2$), and exponential power with a density proportional to $\exp(-|z|^6)$ ($\kappa = -1$). All 5 densities are standardized to zero mean and unit variance. The sample size is $n = 128$. Five α -distance covariance tests with

TABLE 2.9. Empirical Type I error rates of *HSIC* for meta-Gaussian distributions. The kernel is Gaussian with scaling set at the median of distances. The null distribution of *HSIC* is approximated using a gamma distribution as described in Gretton et al. (2008) and a Pearson type III distribution.

Gamma approximation											
n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$	n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$
15	2	3	0.036	0.055	0.063	50	2	3	0.047	0.060	0.061
	5	5	0.017	0.023	0.037		5	5	0.046	0.057	0.050
	10	15	0	0	0.007		10	15	0.020	0.022	0.020
	50	50	0	0	0		50	50	0.001	0	0
25	2	3	0.051	0.056	0.047	100	2	3	0.041	0.054	0.061
	5	5	0.038	0.035	0.038		5	5	0.057	0.052	0.050
	10	15	0.007	0.008	0.007		10	15	0.042	0.046	0.038
	50	50	0	0	0		50	50	0.004	0.006	0.001

Pearson type III approximation											
n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$	n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$
15	2	3	0.032	0.044	0.048	50	2	3	0.045	0.040	0.050
	5	5	0.060	0.062	0.045		5	5	0.066	0.075	0.057
	10	15	0.058	0.040	0.056		10	15	0.051	0.052	0.054
	50	50	0.059	0.058	0.062		50	50	0.052	0.047	0.055
25	2	3	0.050	0.042	0.053	100	2	3	0.046	0.048	0.049
	5	5	0.054	0.060	0.047		5	5	0.047	0.046	0.045
	10	15	0.050	0.054	0.051		10	15	0.033	0.039	0.044
	50	50	0.059	0.059	0.055		50	50	0.057	0.055	0.050

indices $\alpha = 1/3, 2/3, 1, 4/3, 5/3$ were compared to two *HSIC* tests of index $1/3$: $\mathcal{H}_n^{2(1/3)}$ with very small scale parameters $\beta_j = 10^{-5}/\text{med}_{k<l}|Z_k^{(j)} - Z_l^{(j)}|$ and $\mathcal{H}_{n,adap}^{2(1/3)}$ with adaptive selection. For the case $\theta = 0$, Figure 2.1 shows that all the tests reach the nominal level of 5%. Sejdinovic et al. (2013) compared *HSIC* with a Gaussian kernel with scale parameter (2.10) and found that it has poor power relative to $\mathcal{V}_n^{2(1/3)}$. However, from Corollary 2.1, $\mathcal{H}_n^{2(1/3)}$ has the same power as $\mathcal{V}_n^{2(1/3)}$, see the left panel of Figure 2.1. However, $\mathcal{H}_{n,adap}^{2(1/3)}$ is more powerful than $\mathcal{V}_n^{2(1/3)}$. Empirical powers are computed from 1000 tests at nominal level 5%.

Another example borrowed from Sejdinovic et al. (2013) compared two *HSIC* tests of index $1/6$ to distance covariance tests of index $1/6, 1/3, 2/3$, and 1 on sinusoidally dependent data in dimensions $d_1 = d_2 = 1$. The sample size is $n = 512$. The density is $g(z_1, z_2) \propto 1 + \sin(lz_1) \sin(lz_2)$, (z_1, z_2) in $(-\pi, \pi)^2$, where $l = 1, 2, \dots$ is the frequency. In Figure 2.1, the best distance covariance test has index $1/6$ and is matched by the *HSIC* test $\mathcal{H}_n^{2(1/6)}$ of the same index with small scale parameters $\beta_j = 10^{-8}/\text{med}_{k<l}|Z_k^{(j)} - Z_l^{(j)}|$. However, the most powerful test

TABLE 2.10. Empirical Type I error rates of $HSIC$ for meta-Student distributions with 2 degrees of freedom. The kernel is Gaussian with scaling set at the median of distances. The null distribution of $HSIC$ is approximated using a gamma distribution as described in Gretton et al. (2008) and a Pearson type III distribution.

Gamma approximation											
n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$	n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$
15	2	3	0.089	0.048	0.100	50	2	3	0.050	0.043	0.075
	5	5	0.060	0.020	0.061		5	5	0.062	0.042	0.055
	10	15	0.102	0.002	0.135		10	15	0.105	0.019	0.110
	50	50	0.096	0.002	0.148		50	50	0.115	0.042	0.109
25	2	3	0.074	0.044	0.088	100	2	3	0.061	0.053	0.065
	5	5	0.068	0.020	0.069		5	5	0.063	0.059	0.079
	10	15	0.108	0.013	0.147		10	15	0.078	0.038	0.082
	50	50	0.125	0.022	0.111		50	50	0.095	0.048	0.090

Pearson type III approximation											
n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$	n	d_1	d_2	Identity	$\text{sgn}(z) z ^{1/2}$	$\text{sgn}(z) z ^{3/2}$
15	2	3	0.064	0.055	0.057	50	2	3	0.043	0.040	0.052
	5	5	0.053	0.050	0.054		5	5	0.042	0.055	0.038
	10	15	0.054	0.046	0.050		10	15	0.037	0.037	0.047
	50	50	0.053	0.049	0.052		50	50	0.045	0.047	0.042
25	2	3	0.044	0.047	0.061	100	2	3	0.044	0.045	0.048
	5	5	0.062	0.046	0.058		5	5	0.039	0.054	0.049
	10	15	0.060	0.062	0.063		10	15	0.046	0.048	0.050
	50	50	0.064	0.055	0.068		50	50	0.061	0.053	0.066

is $\mathcal{H}_{n,adap}^{2(1/6)}$ which improves on $\mathcal{V}_n^{2(1/6)}$. The model used here does not include the independence case for any values of l . Type I error rates of the tests are based on two independent uniform variables on $(-\pi, \pi)$. Powers and type I error rates are computed from 1000 tests at nominal level 5%. Table 2.11 shows that all the distance covariance and $HSIC$ tests considered reach the nominal level.

TABLE 2.11. Empirical type I error rates of four distance covariance tests and two $HSIC$ tests for the uniform independent variables. The sample size is $n = 512$. The number of replications is 1000.

Statistics	$\mathcal{V}_n^{2(1/6)}$	$\mathcal{V}_n^{2(1/3)}$	$\mathcal{V}_n^{2(2/3)}$	\mathcal{V}_n^2	$\mathcal{H}_n^{2(1/6)}$	$\mathcal{H}_{n,adap}^{2(1/6)}$
Type I error rates	0.042	0.044	0.047	0.052	0.042	0.044

A last example is similar to the simulation of Kojadinovic and Holmes (2009). Empirical type I error rates and powers are computed from 1000 tests at nominal

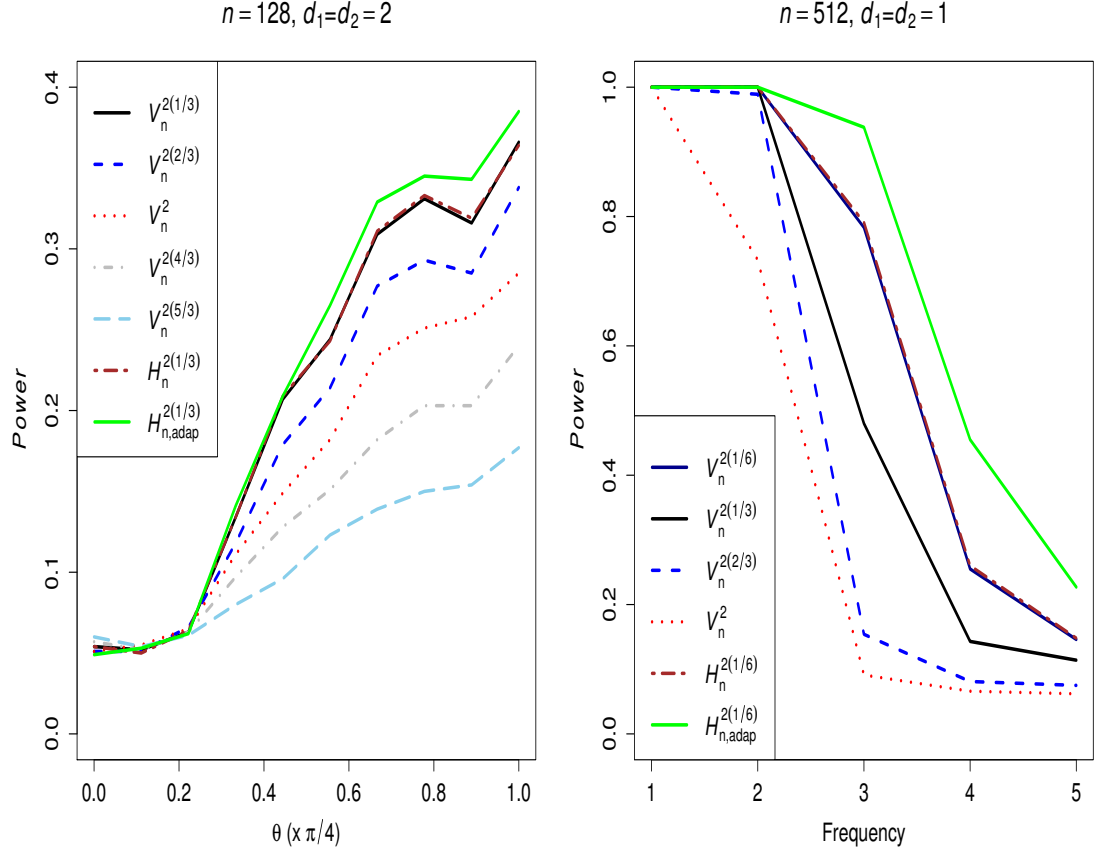


FIGURE 2.1. Left panel: Power of five distance covariance tests and two *HSIC* tests for the ICA benchmark data. Right panel: Power of four distance covariance tests and two *HSIC* tests for the sinusoidally dependent data.

level 5%. The sample size is $n = 100$. Computation of empirical type I error rates used two independent Student copulas of dimensions $d_1 = d_2 = 5$ with 2 degrees of freedom. Evaluation of powers used simulated data from a Student copula of dimension 10 ($d_1 = d_2 = 5$) with 2 degrees of freedom. All variables are then uniformly distributed on $(0, 1)$. The correlation matrix is constructed as follows. A partial correlation matrix is first constructed of the form

$$P = \begin{pmatrix} (1 - \rho_w)I_5 + \rho_w J_5 J_5' & \rho_b J_5 J_5' \\ \rho_b J_5 J_5' & (1 - \rho_w)I_5 + \rho_w J_5 J_5' \end{pmatrix},$$

where J_5 is the vector of ones. The notations w and b stand for within and between, respectively. Then, the one-to-one transformation of Joe (2006) between partial correlations and simple correlations is applied to P to yield a correlation matrix R . The resulting matrix R is positive definite for all ρ_w and ρ_b in $(-1, 1)$, although P may not be positive definite. The values of ρ_w are: 0 and 0.5. For

each value of ρ_w , we consider the value of ρ_b : 0, 0.1, 0.2, 0.3 and 0.4. Four distance covariance tests of index 1/6, 1/3, 2/3, and 1 were compared to two *HSIC* tests of index 1/6: $\mathcal{H}_n^{2(1/6)}$ with very small scale parameters $\beta_j = 10^{-8}/\text{med}_{k<l}|Z_k^{(j)} - Z_l^{(j)}|$ and $\mathcal{H}_{n,\text{adap}}^{2(1/6)}$ with adaptive selection. Table 2.12 shows that all the distance covariance tests and *HSIC* tests reach the nominal level of 5% with the intra correlation values $\rho_w = 0$ and $\rho_w = 0.5$. Figure 2.2 shows that the best distance covariance test has index 1/6 and it is matched by $\mathcal{H}_n^{2(1/6)}$. Again, the test $\mathcal{H}_{n,\text{adap}}^{2(1/6)}$ with adaptive selection has the best power.

TABLE 2.12. Empirical type I error rates simulated from 1000 tests at 5% significance level. Four distance covariance tests and two *HSIC* tests are considered. Simulations are based on two independent Student copulas with two degrees of freedom. The sample size is $n = 100$ and the dimensions are $d_1 = d_2 = 5$.

$\rho_w = 0$						
Statistics	$\mathcal{V}_n^{2(1/6)}$	$\mathcal{V}_n^{2(1/3)}$	$\mathcal{V}_n^{2(2/3)}$	\mathcal{V}_n^2	$\mathcal{H}_n^{2(1/6)}$	$\mathcal{H}_{n,\text{adap}}^{2(1/6)}$
Type I error rates	0.055	0.052	0.054	0.052	0.055	0.057
$\rho_w = 0.5$						
Statistics	$\mathcal{V}_n^{2(1/6)}$	$\mathcal{V}_n^{2(1/3)}$	$\mathcal{V}_n^{2(2/3)}$	\mathcal{V}_n^2	$\mathcal{H}_n^{2(1/6)}$	$\mathcal{H}_{n,\text{adap}}^{2(1/6)}$
Type I error rates	0.043	0.041	0.044	0.047	0.043	0.048

In any case, one can always find a characteristic kernel such that *HSIC* has, for all practical purposes, the same power as any α -distance covariance test. This holds although distance covariance *has not* a strict equivalence to a kernel-based measure of dependence as shown in Sejdinovic et al. (2013, p. 2279). They state informally that the exponent in the α -distance covariance plays a similar role as the bandwidth of the Gaussian kernel. It would be preferable to say that it plays the same role as the index α of the characteristic kernel of a stable distribution. The same performance is then obtained by selecting sufficiently small scale parameters. Moreover, in some cases as shown on the previous examples, improved performance of *HSIC* can be produced by using the proposed adaptive selection.

2.7. CONCLUSION

In the class of tests represented as an inner product of two doubly-centered square matrices, it was found that distance covariance and *HSIC* are very competitive in terms of power. Their p -values can be approximated accurately in large dimensions by the Pearson type III approximation which is fast since it does not require doing any permutation as opposed to randomization tests. The *HSIC*

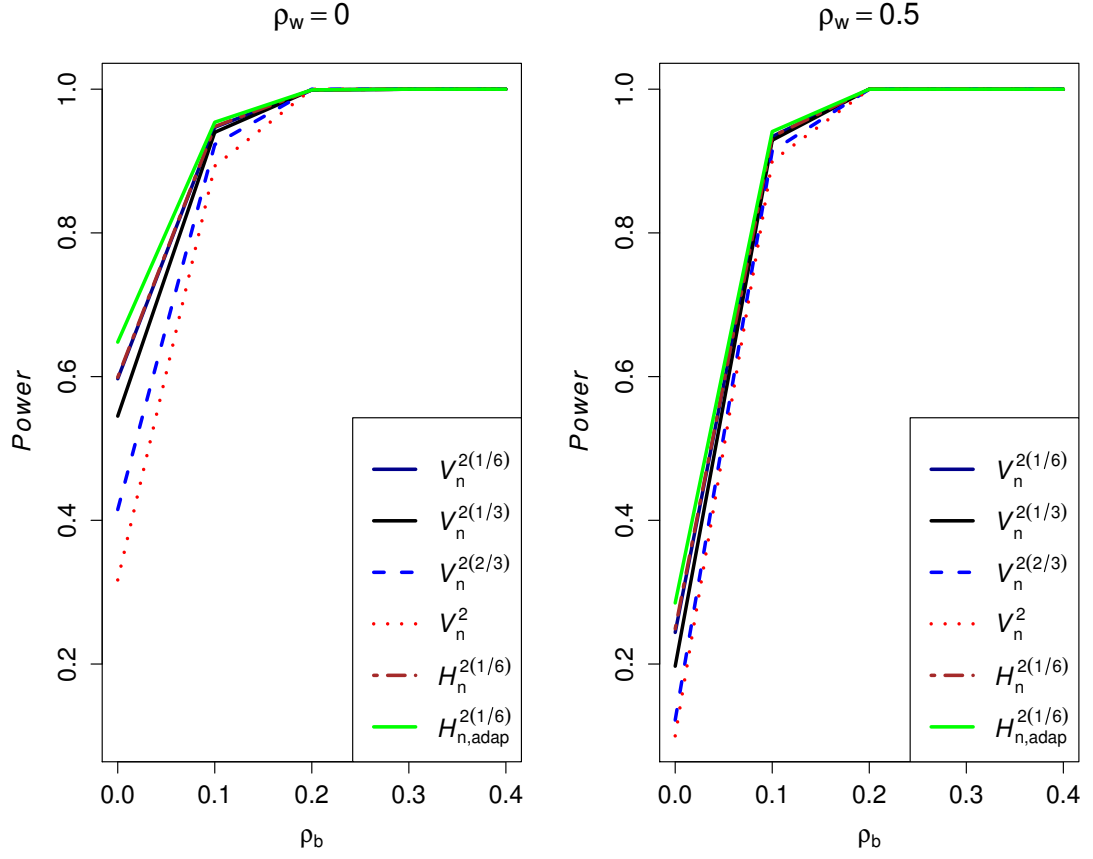


FIGURE 2.2. Power of four distance covariance tests and two *HSIC* tests for data sets following a Student copula with two degrees of freedom. The sample size is $n = 100$ and dimensions are $d_1 = d_2 = 5$.

test with the characteristic kernel of a stable distribution with a sufficiently small scale parameter performs just like a distance covariance test. Both tests require the selection of an index α . However, distance covariance are simpler to compute and do not require the selection of a scale parameter. This may also be a backlash since some scale parameters of *HSIC*, as the new selection method (2.9), may yield higher powers than that of distance covariance.

Finally, the generalization to the problem of testing for mutual independence, not only pairwise independence, between more than two vectors using distance covariance, or *HSIC* tests, is worthy of future research. It would find valuable applications in *nonlinear* Independent Component Analysis (ICA) since mutual independence is no longer equivalent to pairwise independence (Comon, 1994).

BIBLIOGRAPHY

- F.R. Bach and M.I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.
- R. Beran, M. Bilodeau, and P. Lafaye de Micheaux. Nonparametric tests of independence between random vectors. *Journal of Multivariate Analysis*, 98(9):1805–1824, 2007.
- R. Cl  roux and G. Ducharme. Vector correlation for elliptical distributions. *Communication in Statistics-Theory and methods*, 18(4):1441–1454, 1989.
- R. Cl  roux, A. Lazraq, and Y. Lepage. Vector correlation based on ranks and a nonparametric test of no association between vectors. *Communication in Statistics-Theory and methods*, 24(3):713–733, 1995.
- P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.
- L. Cope. Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1279–1281, 2009.
- S. Cs  rg   and C.R. Heathcote. Testing for symmetry. *Biometrika*, 74(1):177–184, 1987.
- J.A.F. Diniz-Filho, T.N. Soares, J.S. Lima, R. Dobrovolski, V.L. Landeiro, M.P. de C. Telles, T.F. Rangel, and L.M. Bini. Mantel test in population genetics. *Genetics and Molecular Biology*, 36(4):475–485, 2013.
- S. Dray and A.B. Dufour. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1–20, 2007.
- Y. Escoufier. Le traitement des variables vectorielles. *Biometrics*, 29(4):751–760, 1973.
- X. Fernique. R  gularit   de processus gaussiens. *Inventiones Mathematicae*, 12(4):304–320, 1971.
- A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419–433, 1993.
- K. Fukumizu, A. Gretton, X. Sun, and B. Sch  lkopf. Kernel measures of conditional dependence. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 489–496. Curran Associates, Inc., 2008.
- C. Genest and B. R  millard. Tests of independence or randomness based on the empirical copula process. *Test*, 13(2):335–370, 2005.
- A. Gretton, O. Bousquet, A. Smola, and B. Sch  lkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In S. Jain, H. Simon, and E. Tomita,

- editors, *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2005.
- A. Gretton, F. Kenji, C.-H. Teo, L. Song, B. Schölkopf, and A. Smola. A kernel statistical test of independence. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 585–592. Curran Associates, Inc., 2008.
- A. Gretton, K. Fukumizu, and B.K. Sriperumbudur. Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1285–1294, 12 2009.
- F. Husson, J. Josse, S. Le, and J. Mazet. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*, 2015. URL <http://CRAN.R-project.org/package=FactoMineR>. R package version 1.31.3.
- H. Joe. Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10), 2006.
- J. Josse and S. Holmes. Tests of independence and beyond. *arXiv:1307.7383v3*, 2014.
- J. Josse, J. Pagès, and F. Husson. Testing the significance of the rv coefficient. *Computational Statistics & Data Analysis*, 53(1):82–91, 2008.
- F. Kazi-Aoual, S. Hitier, R. Sabatier, and J.-D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics & Data Analysis*, 20(6):643–656, 1995.
- I. Kojadinovic and M. Holmes. Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *Journal of Multivariate Analysis*, 100(6):1137–1154, 2009.
- I. Kojadinovic and J. Yan. Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9): 1–20, 2010.
- R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5): 3284–3305, 2013.
- N. Mantel. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1):209–220, 1967.
- C.-D. Mayer, J. Lorent, and G.W. Horgan. Exploratory analysis of multiple omics datasets using the adjusted rv coefficient. *Statistical Applications in Genetics and Molecular Biology*, 10(1):1–27, 2011.
- C. Minas, E. Curry, and G. Montana. A distance-based test of association between paired heterogeneous genomic data. *Bioinformatics*, 29(20):2555–2563, 2013.
- M. Omelka and Š. Hudecová. A comparison of the Mantel test with a generalised distance covariance test. *Environmetrics*, 24(7):449–460, 2013.

- M.L. Rizzo and G.J. Székely. *energy: E-statistics (energy statistics)*, 2014. URL <http://CRAN.R-project.org/package=energy>. R package version 1.6.2.
- P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods: The rv- coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25(3):257–265, 1976.
- P. Robert, R. Cléroux, and N. Ranger. Some results on vector correlation. *Computational Statistics & Data Analysis*, 3:25–32, 1985.
- D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291, 2013.
- M.S. Srivastava and N. Reid. Testing the structure of the covariance matrix with fewer observations than the dimension. *Journal of Multivariate Analysis*, 112: 156–171, 2012.
- W. Stute. The oscillation behavior of empirical processes: The multivariate case. *The Annals of Probability*, 12(2):361–379, 1984.
- G.J. Székely and M.L. Rizzo. Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1236–1265, 2009a.
- G.J. Székely and M.L. Rizzo. Rejoinder: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1303–1308, 2009b.
- G.J. Székely, M.L. Rizzo, and N.K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

Chapter 3

APPROXIMATIONS TO PERMUTATION TESTS OF MUTUAL OR SERIAL INDEPENDENCE OF RANDOM VECTORS

Cet article a été soumis à la revue *Journal of Machine Learning Research*.

Les principales contributions de Aurélien Guetsop Nangue à cet article sont présentées.

- Conduite d’une importante revue de la littérature.
- Démonstration des théorèmes 3.5, 3.6 et 3.9.
- Production globale des sections 3.7 et 3.8.
- Conception, écriture et validation des programmes R.
- Conduite des simulations et des applications.
- Rédaction d’une partie de l’article.

Approximations to permutation tests of mutual or serial independence of random vectors

Martin Bilodeau and Aurélien Guetsop Nangue

Département de mathématiques et de statistique,

Université de Montréal,

C.P. 6128, Succursale A,

Montréal, Canada H3C 3J7

Abstract

The problem of testing mutual independence between many random vectors is addressed. The closely related problem of testing serial independence of a multivariate stationary sequence is also considered. The Möbius transformation of characteristic functions is used to characterize independence. A generalization to p vectors of the α -distance covariance test and the Hilbert-Schmidt independence criterion (*HSIC*) test with the characteristic kernel of a stable probability distribution of index α is obtained. It is shown that an *HSIC* test with sufficiently small scale parameters is equivalent to an α -distance covariance test. Weak convergence of the *HSIC* test is established. A very fast and accurate computation of p -values uses the Pearson type III approximation which successfully approaches the exact permutation distribution of the tests. This approximation relies on the exact first three moments of the permutation distribution of any test which can be expressed as the sum of all elements of a componentwise product of p doubly-centered matrices. The α -distance covariance test and the *HSIC* test are both of this form. A new selection method is proposed for the scale parameter of the characteristic kernel of the *HSIC* test. It is shown in a simulation that this adaptive *HSIC* test has higher power than the α -distance covariance test when data are generated from a Student copula. Applications are given to environmental and financial data.

Keywords: Characteristic function, distance covariance, Hilbert-Schmidt independence criterion, mutual independence, Möbius transformation, permutation test, quadratic distance, serial independence.

3.1. INTRODUCTION

The problem of testing for independence between p components of a random vector has attracted considerable attention in statistics. Many nonparametric procedures exist in the literature. A natural approach is to consider a functional of the difference between the empirical joint distribution and the product of the empirical marginal distributions. This same approach can also use empirical characteristic functions. When the functional of the difference is above a certain threshold, the components are declared dependent. Csörgő (1985), Kankainen (1995), Sejdinovic et al. (2013b) and Fan et al. (2015) considered mutual tests of independence based on empirical characteristic functions. However, when dependence is declared, it is not possible to identify, with their proposed tests, subsets of variables responsible for the dependence. This limitation is similar to that of a global F -test in an analysis of variance model with one fixed factor, as opposed to multiple comparisons procedures. Or, that of a global chi-square test of independence in a multi-way contingency table, as opposed to log-linear models with interaction terms. For tests of independence, a useful method is the Möbius transformation.

The Möbius transformation has a long history in statistics. The Möbius transformation of distribution functions was first proposed in Blum et al. (1961) for $p = 3$. The general case was treated in Deheuvels (1981), Ghoudi et al. (2001), Genest and Rémillard (2005), Kojadinovic and Holmes (2009), Kojadinovic and Yan (2011), and Duchesne et al. (2012). It can also be defined with characteristic functions as in Bilodeau and Lafaye de Micheaux (2005), with half-space probabilities as in Beran et al. (2007), or with cell probabilities in a contingency table as in Bilodeau and Lafaye de Micheaux (2009). The first appearance of a Möbius transformation, although not stated explicitly, goes back to the work of Lancaster (1951) on contingency tables as explained in Bilodeau and Lafaye de Micheaux (2009). The machine learning community (Sejdinovic et al., 2013a) proposed kernel nonparametric tests for Lancaster three-variable interaction. This test is in fact a test based on the empirical version of the Möbius transformation of the characteristic function when $p = 3$. The general Möbius transformation considered in this thesis can be used to build tests for general interactions of any order, as well as tests of mutual and serial independence.

The Chapter is organized as follows. Section 3.2 introduces the Möbius transformation of characteristic functions. It presents a characterization of the mutual independence between p random vectors by the Möbius transformation. In Section 3.3, new tests based on the Möbius transformation of empirical characteristic

functions are introduced. They generalize the Hilbert-Schmidt independence criterion test (*HSIC*) (Gretton et al., 2005, 2009) and the distance covariance test (Székely et al., 2007) to the case $p > 2$. The new tests have a common form as a sum of elements of a componentwise product of p doubly-centered matrices. An equivalence is established between an *HSIC* test with infinitesimal scale parameters and a distance covariance test. The weak convergence of the empirical process based on the Möbius transformation is proved. The weak convergence of the *HSIC* test is also established. A difficulty encountered in establishing the weak convergence of the distance covariance test is described.

Rather than relying on the limiting distribution to conduct tests, Section 3.4 proposes the Pearson type III approximation to permutation tests as a very fast and more accurate approximation. A major contribution provides expressions for the exact first three moments of the permutation distribution. The elaborate expression for the third moment is relegated to Appendix B. For *HSIC* tests with characteristic kernels, a novel and promising method for the adaptive selection of scale parameters is proposed. It selects the scale parameters which maximize the variance of the permutation distribution. This optimization can be easily achieved with a Newton-type algorithm with numerical derivatives. In Section 3.5, p -values obtained by the Pearson type III approximation for all possible subsets of components are combined à la Fisher to obtain a global test of mutual independence. Section 3.6 adapts all the results described for the mutual independence situation to the problem of testing for the serial independence of a stationary sequence.

Section 3.7 adapts the dependogram, a graphical device of Genest and Rémillard (2005), to *HSIC* and distance covariance tests. Two examples with simulated continuous data borrowed from Kojadinovic and Holmes (2009) are revisited. They illustrate the computational speed of the tests and the greater power of the *HSIC* test with adaptive selection. Another example considers simulated data consisting of a sequence of binary digits exhibiting a serial dependence at lag 3 detected successfully by a distance covariance test and an adaptive *HSIC* test.

Finally, Section 3.8 contains an application to real data on variables related to air temperature, soil temperature, humidity, wind, and evaporation. *HSIC* or distance covariance tests should be preferred to the Gaussian likelihood ratio test since a multivariate Gaussian model is rejected by the test of Henze and Zirkler (1990). Another application finds significant serial dependencies in the increasing daily rates of the S&P/TSX composite, S&P500 and DOW JONES

indices ranging from January 2, 2014 to March 2, 2016. All proofs are given in Appendices A and B.

3.2. MÖBIUS TRANSFORMATION

The Möbius transformation of characteristic functions is a powerful tool for the characterization of mutual independence between p random vectors $Z^{(1)}, \dots, Z^{(p)}$. The dimension of the vector $Z^{(j)}$ is d_j , for $j = 1, \dots, p$. Let f be the joint characteristic function of these p vectors, and let $f^{(j)}$ be the marginal characteristic function of $Z^{(j)}$. Mutual independence is characterized by the factorization

$$f(t^{(1)}, \dots, t^{(p)}) = \prod_{j=1}^p f^{(j)}(t^{(j)}),$$

for all $t^{(1)}, \dots, t^{(p)}$. It may also be characterized by the Möbius transformation which is defined as follows. Let \mathcal{I}_p be the family of subsets B of $\{1, \dots, p\}$ of cardinality $|B| > 1$. The set \mathcal{I}_p has $2^p - p - 1$ elements since the empty set is excluded, as well as all p singletons. For any $B \in \mathcal{I}_p$ and any $t^{(1)}, \dots, t^{(p)}$, define $t^{(B)} = \{t^{(j)} : j \in B\}$. Similarly, $Z^{(B)} = \{Z^{(j)} : j \in B\}$, and $f^{(B)}$ is the joint characteristic function of $Z^{(B)}$. The Möbius transformation of the characteristic function f for the set $B \in \mathcal{I}_p$ is given by

$$\mu_B(t^{(B)}) = \sum_{C \subset B} (-1)^{|B \setminus C|} f^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f^{(j)}(t^{(j)}).$$

The following characterization holds: $Z^{(1)}, \dots, Z^{(p)}$ are mutually independent if and only if, $\mu_B(t^{(B)}) = 0$, for all $B \in \mathcal{I}_p$, and all vectors $t^{(B)}$. A proof by induction of this characterization using distribution functions is given in Ghoudi et al. (2001) and is immediately applicable to characteristic functions.

3.3. DEPENDENCE STATISTICS

Consider $(Z_k^{(1)}, \dots, Z_k^{(p)})$, $k = 1, \dots, n$, an independent and identically distributed sample of size n . The processes corresponding to μ_B are defined as

$$R_{nB}(t^{(B)}) = \sqrt{n} \sum_{C \subset B} (-1)^{|B \setminus C|} f_n^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f_n^{(j)}(t^{(j)}), \quad (3.1)$$

where

$$f_n^{(C)}(t^{(C)}) = \frac{1}{n} \sum_{k=1}^n e^{i \langle t^{(C)}, Z_k^{(C)} \rangle}$$

is the empirical characteristic function. When $C = \{j\}$ is a singleton, the notation used for $f_n^{(C)}$ is simply $f_n^{(j)}$. The dependence statistic for the subset B is now

defined as the functional

$$T_{nB} = \int |R_{nB}(t^{(B)})|^2 dw_B, \quad (3.2)$$

where dw_B is a product measure

$$dw_B = \prod_{j \in B} dw^{(j)}(t^{(j)}).$$

The evaluation of this integral is facilitated using another representation of the process. First, recall the multinomial formula (Ghoudi et al., 2001)

$$\sum_{C \subset B} \left(\prod_{i \in C} u^{(i)} \right) \left(\prod_{j \in B \setminus C} v^{(j)} \right) = \prod_{i \in B} (u^{(i)} + v^{(i)}).$$

Then, the empirical process (3.1) can be written as

$$R_{nB}(t^{(B)}) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \prod_{j \in B} \left[e^{i \langle t^{(j)}, Z_k^{(j)} \rangle} - f_n^{(j)}(t^{(j)}) \right]. \quad (3.3)$$

The result in Equation (3.3) is obtained by replacing the expression $f_n^{(C)}(t^{(C)}) = (1/n) \sum_{k=1}^n e^{i \langle t^{(C)}, Z_k^{(C)} \rangle}$ in Equation (3.1) and by applying the multinomial formula. The representation given by (3.3) allows the integral (3.2) to be evaluated explicitly in some cases and simplifies the proofs of theorems to come. Two important cases are now presented.

3.3.1. Hilbert-Schmidt independence criterion

The measure $dw_B = \prod_{j \in B} dG^{(j)}(t^{(j)})$ is a product of probability measures. Let $\varphi^{(j)}$ be the characteristic function of $G^{(j)}$.

Theorem 3.1. *For $k, l = 1, \dots, n$ and $j = 1, \dots, p$, let*

$$a_{kl}^{(j)} = \varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)}). \quad (3.4)$$

Then, for any $B \in \mathcal{I}_p$, the dependence statistic T_{nB} is given by

$$T_{nB} = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \prod_{j \in B} A_{kl}^{(j)}, \quad (3.5)$$

where $A_{kl}^{(j)} = a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}$, $j = 1, \dots, p$ and $k, l = 1, \dots, n$ and

$$\bar{a}_{k.}^{(j)} = \frac{1}{n} \sum_{l=1}^n a_{kl}^{(j)}, \quad \bar{a}_{.l}^{(j)} = \frac{1}{n} \sum_{k=1}^n a_{kl}^{(j)}, \quad \bar{a}_{..}^{(j)} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}^{(j)}.$$

By definition, matrices $A^{(j)}$, $j = 1, \dots, p$, are doubly-centered, *i.e.* rows and columns of these matrices sum up to zero. In the special case $p = 2$ of testing the independence between two vectors, the statistic T_{nB} , for $B = \{1, 2\}$, is the

Hilbert-Schmidt independence criterion, or *HSIC*, with characteristic kernels $\varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)})$, $j = 1, 2$.

An important special case is when $G^{(j)}$ is the stable distribution of index $\alpha \in (0, 2]$ with scale parameter $\beta_j > 0$. Then, the characteristic kernel is

$$a_{kl}^{(j)} = e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha}, \quad (3.6)$$

where $|\cdot|_{d_j}$ is the Euclidian norm in dimension d_j . The corresponding dependence statistic is then denoted $T_{nB} = \mathcal{H}_{nB}^{2(\alpha)}$. The case $\alpha = 2$ is the Gaussian kernel, and $\alpha = 1$ is the Cauchy kernel, often referred to as the Laplace kernel in machine learning (Gretton et al., 2005, 2009) because of its similarity to a Laplace density in dimension one.

3.3.2. Distance covariance

The measure $dw_B = \prod_{j \in B} dw^{(j)}(t^{(j)})$ is a product of non integrable measures. The measure $dw^{(j)}$ is defined as

$$dw^{(j)}(t^{(j)}) = [C(d_j, \alpha) |t^{(j)}|_{d_j}^{d_j + \alpha}]^{-1},$$

with the normalizing constant

$$C(d, \alpha) = 2\pi^{d/2} \Gamma(1 - \alpha/2) / [\alpha 2^\alpha \Gamma((d + \alpha)/2)].$$

A similar representation holds.

Theorem 3.2. *For $k, l = 1, \dots, n$ and $j = 1, \dots, p$, let*

$$a_{kl}^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha. \quad (3.7)$$

Then, for any $B \in \mathcal{I}_p$, the dependence statistic T_{nB} has the same form as in Theorem 3.1.

The corresponding dependence statistic is then denoted $T_{nB} = \mathcal{V}_{nB}^{2(\alpha)}$, which is the usual notation for distance covariance of index α (Székely et al., 2007). Again in the special case $p = 2$, the statistic T_{nB} reduces to the distance covariance of index α . A very special case is when $p = 2$ and $\alpha = 2$. In this case, $\mathcal{V}_{nB}^{2(2)}$ is the numerator of the *RV* coefficient of Escoufier (1973) as noticed by Székely et al. (2007). The following limit establishes that $\mathcal{V}_{nB}^{2(\alpha)}$ is, for all practical purpose, equivalent to $\mathcal{H}_{nB}^{2(\alpha)}$ when scale parameters β_j , $j \in B$, are sufficiently small:

$$\lim_{\beta_j \rightarrow 0, \forall j \in B} \mathcal{H}_{nB}^{2(\alpha)} / \prod_{j \in B} \beta_j^\alpha = \mathcal{V}_{nB}^{2(\alpha)}. \quad (3.8)$$

This result, proved in Appendix A, implies that $\mathcal{H}_{nB}^{2(\alpha)}$, with sufficiently small scale parameters, will have a power function indistinguishable from that of $\mathcal{V}_{nB}^{2(\alpha)}$.

As simple as it may seem, this equivalence, for the simplest case $p = 2$, has gone unnoticed in the discussions of distance covariance (Székely and Rizzo, 2009; Gretton et al., 2009). In Section 8.2 of Sejdinovic et al. (2013b), the *HSIC* test $\mathcal{H}_n^{2(2)}$ with Gaussian kernels with scale parameters set at the inverse of median of interpoint distances is compared to distance covariance tests of varying index α . It was found in the independent component analysis benchmark example that $\mathcal{V}_n^{2(1/3)}$ is more powerful than $\mathcal{H}_n^{2(2)}$. From (3.8), the *HSIC* test $\mathcal{H}_n^{2(1/3)}$, with characteristic kernels of the stable distribution of index $\alpha = 1/3$ and very small scale parameters, would have a power function indistinguishable from that of $\mathcal{V}_n^{2(1/3)}$. In another example with sinusoidally dependent data, the *HSIC* test $\mathcal{H}_n^{2(2)}$ has a very poor power function compared to $\mathcal{V}_n^{2(1/6)}$. Sejdinovic et al. (2013b) explained: “the exponent in the distance-induced kernel plays a similar role as the bandwidth of the Gaussian kernel, and smaller exponents are able to detect dependencies at smaller lengthscales. Poor performance of the Gaussian kernel with median bandwidth in this example is a consequence of the mismatch between the overall lengthscale of the marginal distributions (captured by the median inter-point distances) and the lengthscales at which dependencies are present”. In fact, the exponent in the distance-induced kernel of a distance covariance test plays the same role as the index of the characteristic kernel of a stable distribution in an *HSIC* test. Indistinguishable power functions can be obtained by choosing sufficiently small scale parameters in the characteristic kernel. This all means that *HSIC* with sufficiently small scale parameters always match distance covariance in terms of power. But *HSIC* with scale parameters appropriately selected may, in some cases, improve on distance covariance.

3.3.3. Asymptotic distribution

Empirical processes as in (3.1) have been recently very useful at tackling problems related to mutual independence because of the simplicity of the asymptotic distribution. Let $d_B = \sum_{j \in B} d_j$. Each process R_{nB} is defined on the space $C(\mathbb{R}^{d_B}, \mathbb{C})$ of complex-valued continuous functions defined on \mathbb{R}^{d_B} . Proofs are given in Appendix A.

Theorem 3.3. *If $Z^{(1)}, \dots, Z^{(p)}$ are mutually independent, then the collection of processes $\{R_{nB}, B \in \mathcal{I}_p\}$ converges in $C(\mathbb{R}^{d_B}, \mathbb{C})$ to independent zero mean complex Gaussian processes R_B with $\bar{R}_B(t^{(B)}) = R_B(-t^{(B)})$ and complex covariance functions*

$$\mathbb{E} [R_B(t^{(B)}) \bar{R}_B(s^{(B)})] = \prod_{j \in B} [f^{(j)}(t^{(j)} - s^{(j)}) - f^{(j)}(t^{(j)}) f^{(j)}(-s^{(j)})]. \quad (3.9)$$

The convergence of functionals (3.2) also holds even though it is not defined on the whole space $C(\mathbb{R}^{d_B}, \mathbb{C})$, but only on the space of squared integrable functions. In the next theorem, the asymptotic distribution of $T_{nB} = \mathcal{H}_{nB}^{2(\alpha)}$ is described for any $\alpha \in (0, 2]$.

Theorem 3.4. *Without loss of generality, let $B = \{1, \dots, k\}$. If $Z^{(1)}, \dots, Z^{(p)}$ are mutually independent, then $T_{nB} \rightsquigarrow T_B$ for each $B \in \mathcal{I}_p$, with T_B being distributed as*

$$\sum_{i_1=1}^{\infty} \dots \sum_{i_k=1}^{\infty} \lambda_{i_1}^{(1)} \dots \lambda_{i_k}^{(k)} Z_{i_1 \dots i_k}^2, \quad (3.10)$$

where $Z_{i_1 \dots i_k}$ are independent standard normal variables, and for $j = 1, \dots, k$, $\lambda_1^{(j)}, \lambda_2^{(j)}, \dots$ are eigenvalues depending only on the distribution of $Z^{(j)}$ and the measure dw_j .

The symbol “ \rightsquigarrow ” refers to the weak convergence of processes. Theorem 3.4 holds without moment conditions for $\mathcal{H}_{nB}^{2(\alpha)}$ since dw_B is a probability measure. Theorems 3.3 and 3.4 were proved when $p = 2$ by Zhang et al. (2011).

The same type of results for $p = 2$ have been obtained by Székely et al. (2007) for distance covariance, see also Lyons (2013, Corollary 2.8 and Remark 2.9) for the product structure of eigenvalues. These results are also expected to hold for $p > 2$ under the moment assumption $\mathbb{E}|Z^{(j)}|^\alpha < \infty$ for $j = 1, \dots, p$, although this is not proved here because of the difficulty of handling non integrable measure. In the proof of Theorem 3.4, the generalization of a result of Kellermeier (1980) no longer holds since it is based on Jensen’s inequality valid only for probability measures.

Despite the simple form of this asymptotic distribution, it would be a formidable task to use an approximation based on the computations of these eigenvalues. Moreover, approximations using eigenvalues for the most simple *RV* coefficient of Escoufier (1973) are appropriate only in very large samples (Josse and Holmes, 2014).

Approximations are usually obtained by resampling techniques. The null distribution is simulated by permuting (resample without replacement) the observations $Z_1^{(j)}, \dots, Z_n^{(j)}$, independently for each component. The *permutation test* recomputes the dependence statistics T_{nB} , $B \in \mathcal{I}_p$, for all $(n!)^p$ permutations. Since this is not possible, even for moderate sample sizes, the strategy is to resort to a *randomized test* which consists of using only a large number but still feasible of, say, 1000 permutations. Another strategy is to resample with replacement a large number of times which is a *bootstrap test*.

Tests of mutual independence are often built from the empirical characteristic independence process (Csörgő, 1985; Kankainen, 1995; Sejdinovic et al., 2013a;

Fan et al., 2015),

$$\begin{aligned}\mathcal{J}_n &= n \int \left| f_n(t^{(1)}, \dots, t^{(p)}) - \prod_{j=1}^p f_n^{(j)}(t^{(j)}) \right|^2 \prod_{j=1}^p dw^{(j)}(t^{(j)}) \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \prod_{j=1}^p a_{kl}^{(j)} - \frac{2}{n^p} \sum_{k=1}^n \prod_{j=1}^p \sum_{l=1}^n a_{kl}^{(j)} + \frac{1}{n^{2p-1}} \prod_{j=1}^p \sum_{k=1}^n \sum_{l=1}^n a_{kl}^{(j)}, \quad (3.11)\end{aligned}$$

where $dw^{(1)}, \dots, dw^{(p)}$ are probability measures and $a_{kl}^{(j)}$ is defined in (3.4). Similar tests based on empirical distribution functions have been proposed by (Blum et al., 1961), Cotterill and Csörgő (1982), Cotterill and Csörgő (1985), and Kojadinovic and Holmes (2009). Most of these developments are in the univariate situation, *i.e.* $d_1 = \dots = d_p = 1$, or are applicable only in small dimensions. Estimation of eigenvalues in Fan et al. (2015) necessitates numerical integration in dimension d_j , for each $j = 1, \dots, p$, by Monte Carlo simulations or some cubature rule. Sejdinovic et al. (2013a) proposed to approximate the null distribution by a randomized test.

The asymptotic independence of the processes R_{nB} , $B \in \mathcal{I}_p$, and the product structure of the summands in expression (3.5) for T_{nB} are of paramount importance. The dependence statistics resulting from the Möbius decomposition have advantages over tests of the type \mathcal{J}_n in Equation (3.11).

- (1) The product structure of the summands in (3.5) allows for a fast and accurate Pearson type III approximation based on the exact first three moments which can be used to approach the permutation distribution of T_{nB} . This can be done in high dimensions.
- (2) The p -values of T_{nB} , $B \in \mathcal{I}_p$, can be easily combined à la Fisher (see Section 3.5) to get a test of mutual independence with a given type I global error rate. This is made possible by the asymptotic independence of the processes R_{nB} .
- (3) When Fisher's test rejects the mutual independence hypothesis, subsets $B \in \mathcal{I}_p$ yielding small p -values can be identified as the possible source of dependence.

The first advantage means that tests can be conducted without requiring estimation of eigenvalues obtained by numerically solving integral equations over spaces \mathbb{R}^{d_j} , nor requiring any resampling as for randomized or bootstrap tests. The advantage in item 2 simplifies the analysis. For example, when testing multiple sub-hypotheses, Sejdinovic et al. (2013a) must resort to a sequentially rejection Bonferroni method (Holm, 1979). The decomposition in item 3 bears a

resemblance with a global F test in an analysis of variance followed by multiple comparisons à la Tukey or Scheffé between treatment means.

3.4. A FAST APPROXIMATION TO THE PERMUTATION TEST

The approximation proposed is fast because it does not estimate eigenvalues and does not rely on resampling. The fast approximation uses the exact first three moments of the permutation test in a Pearson type III density. When $p = 2$, the exact first three moments of the permutation test RV were derived by Kazi-Aoual et al. (1995) and Minas et al. (2013, p. 4) noted that the same formula may be used for the permutation test T_{nB} in (3.5) with any doubly-centered matrices $A^{(j)}$, $j = 1, 2$. This led to the Pearson type III approximations of the RV test (Kazi-Aoual et al., 1995; Josse et al., 2008) and of the $HSIC$ test with general kernels (Minas et al., 2013). Expressions for moments are now presented for any p . The exact first two moments of T_{nB} , for any $p \geq 2$, are given in Theorems 3.5 and 3.6. The exact third moment is given in Theorem 3.9 of Appendix B. Its proof is also available in Appendix B.

The set of $n!$ permutations of $1, \dots, n$ is denoted \mathcal{S}_n . For $j = 1, \dots, p$, the notation \sum_{σ_j} is a shorthand for $\sum_{\sigma_j \in \mathcal{S}_n}$. After a permutation $\sigma_j = (\sigma_j(1), \dots, \sigma_j(n))$ of observations $Z_1^{(j)}, \dots, Z_n^{(j)}$, the matrix $A^{(j)}(\sigma_j)$ is obtained by permutation of the rows and corresponding columns of $A^{(j)}$; there is no need to recompute the interpoint distances $|Z_k^{(j)} - Z_l^{(j)}|$. The dependence statistics T_{nB} are all of the generic form $T = \sum_{k,l=1}^n \prod_{j=1}^p A_{kl}^{(j)}$. The exact moment of order s of the permutation test T is

$$\mathbb{E}_P(T^s) = \frac{1}{(n!)^p} \sum_{\sigma_1} \dots \sum_{\sigma_p} \left[\sum_{k,l=1}^n \prod_{j=1}^p A_{kl}^{(j)}(\sigma_j) \right]^s. \quad (3.12)$$

For small values of p and n , the exact moments can be computed by a complete enumeration using (3.12) and used to verify the moment formulas below. The first two moments are expressed using the following notations. Additional notations are introduced in Appendix B for the third moment. For $j = 1, \dots, p$, define the quantities

$$T^{(j)} = \text{tr} [A^{(j)}], \quad T_2^{(j)} = \text{tr} [(A^{(j)})^2], \quad S_2^{(j)} = \sum_k (A_{kk}^{(j)})^2.$$

Theorem 3.5. *The first moment of $T = \sum_{k,l=1}^n \prod_{j=1}^p A_{kl}^{(j)}$, where $A^{(j)}$ is doubly-centered, is given by*

$$(n!)^{p-1} \mathbb{E}_P(T) = [(n-1)!]^{p-1} \prod_{j=1}^p T^{(j)} + [(n-2)!]^{p-1} \prod_{j=1}^p [-T^{(j)}].$$

Theorem 3.6. *The second moment of $T = \sum_{k,l=1}^n \prod_{j=1}^p A_{kl}^{(j)}$, where $A^{(j)}$ is doubly-centered, is given by*

$$\begin{aligned} (n!)^{p-1} \mathbb{E}_P(T^2) &= [(n-1)!]^{p-1} \prod_{j=1}^p S_2^{(j)} + [(n-2)!]^{p-1} \prod_{j=1}^p [(T^{(j)})^2 - S_2^{(j)}] \\ &+ 2[(n-2)!]^{p-1} \prod_{j=1}^p [T_2^{(j)} - S_2^{(j)}] + 4[(n-2)!]^{p-1} \prod_{j=1}^p [-S_2^{(j)}] \\ &+ 4[(n-3)!]^{p-1} \prod_{j=1}^p [-T_2^{(j)} + 2S_2^{(j)}] + 2[(n-3)!]^{p-1} \prod_{j=1}^p [-(T^{(j)})^2 + 2S_2^{(j)}] \\ &+ [(n-4)!]^{p-1} \prod_{j=1}^p [(T^{(j)})^2 + 2T_2^{(j)} - 6S_2^{(j)}]. \end{aligned}$$

The mean ν , variance σ^2 , and skewness γ of the permutation test T can be computed. The Pearson type III approximation for the density of the standardized variable $(T - \nu)/\sigma$ is the shifted-gamma density

$$g(t) = \frac{(2/\gamma)^{4/\gamma^2}}{\Gamma(4/\gamma^2)} \left(\frac{2 + \gamma t}{\gamma} \right)^{(4-\gamma^2)/\gamma^2} e^{-2(2+\gamma t)/\gamma^2}, \quad t > -2/\gamma,$$

with zero mean, unit variance, and skewness γ . The resulting approximation is satisfying only when $\gamma > 0$. For $\gamma < 0$, the approximation is applied to $-T$.

3.4.1. Adaptive scale parameter selection of *HSIC*

Consider the *HSIC* test $\mathcal{H}_{nB}^{2(\alpha)}$ with the characteristic kernel (3.6) of index α . The adaptive scale parameter selection of *HSIC* is a major difficulty which has not been satisfactorily addressed in the literature so far. A new adaptive selection in the special case $p = 2$ was proposed in Chapter 2. The very similar method for the general case $p \geq 2$ is as follows. Without loss of generality, consider the subset $B = \{1, \dots, k\}$. Let $\text{var}_P[n\mathcal{H}_{nB}^{2(\alpha)}(\beta_1, \dots, \beta_k)]$ be the exact variance of the permutation distribution available from Theorems 3.5 and 3.6. The adaptive selection proposed is the solution of the optimization problem,

$$(\hat{\beta}_1, \dots, \hat{\beta}_k) = \arg \sup_{\beta_j > 0, j \in B} \text{var}_P[n\mathcal{H}_{nB}^{2(\alpha)}(\beta_1, \dots, \beta_k)]. \quad (3.13)$$

The motivation for this method was presented in Section 2.5.1 and is the same here. The explicit expression of the exact permutation variance is given by

$$\begin{aligned} &(n!)^{k-1} \text{var}_P[n\mathcal{H}_{nB}^{2(\alpha)}(\beta_1, \dots, \beta_k)] \\ &= [(n-1)!]^{k-1} \prod_{j=1}^k S_2^{(j)} + [(n-2)!]^{k-1} \prod_{j=1}^k [(T^{(j)})^2 - S_2^{(j)}] \end{aligned}$$

$$\begin{aligned}
& +2[(n-2)!]^{k-1} \prod_{j=1}^k [T_2^{(j)} - S_2^{(j)}] + 4[(n-2)!]^{k-1} \prod_{j=1}^k [-S_2^{(j)}] \\
& +4[(n-3)!]^{k-1} \prod_{j=1}^k [-T_2^{(j)} + 2S_2^{(j)}] + 2[(n-3)!]^{k-1} \prod_{j=1}^k [-(T^{(j)})^2 + 2S_2^{(j)}] \\
& +[(n-4)!]^{k-1} \prod_{j=1}^k [(T^{(j)})^2 + 2T_2^{(j)} - 6S_2^{(j)}] \\
& -(n!)^{-k+1} \left[[(n-1)!]^{k-1} \prod_{j=1}^k T^{(j)} + [(n-2)!]^{k-1} \prod_{j=1}^k [-T^{(j)}] \right]^2,
\end{aligned}$$

where for $j = 1, \dots, k$, the quantities $T^{(j)}$, $T_2^{(j)}$ and $S_2^{(j)}$ are computed from the matrix $A^{(j)}$ which is obtained by doubly-centering the matrix $a^{(j)}$ with elements $a_{kl}^{(j)} = e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha}$. Using the above explicit expression of the exact permutation variance as a function of scale parameters, the optimization (3.13) can be rapidly achieved with a Newton-type algorithm with numerical derivatives.

For practical purposes, adaptive selection $(\hat{\beta}_1, \dots, \hat{\beta}_p)$ for the whole set $\{1, \dots, p\}$ are first obtained. Then, for all subsets $B \subset \{1, \dots, p\}$, $B > 1$, adaptive scale parameters selected are those of the corresponding sub-vector of $(\hat{\beta}_1, \dots, \hat{\beta}_p)$. This means that the optimization (3.13) is done only once for the whole set to reduce the computational load.

3.5. COMBINING p -VALUES

A p -value is computed for every dependence statistic T_{nB} , $B \in \mathcal{I}_p$, using the Pearson type III approximation. Denote this p -value by \hat{p}_{nB} . Under the mutual independence hypothesis, the p -values \hat{p}_{nB} , for $B \in \mathcal{I}_p$, are approximated by independent variables uniformly distributed on $(0, 1)$. The quantity $-2 \log \hat{p}_{nB}$ thus has approximately a χ_2^2 null distribution. A test of mutual independence combining these p -values à la Fisher (Littell and Folks, 1971; Genest and Rémillard, 2005) is asymptotically Bahadur optimal (Littell and Folks, 1973). It reports a combined p -value of

$$P \left[\chi_f^2 > -2 \sum_{B \in \mathcal{I}_p} \log \hat{p}_{nB} \right], \text{ where } f = 2(2^p - p - 1).$$

The number of subsets may be too large for some p . In this case, a compromise is to use a test of independence of order m , where $2 \leq m < p$, with p -values

computed as in

$$P \left[\chi_f^2 > -2 \sum_{B \in \mathcal{I}_p, |B| \leq m} \log \hat{p}_{nB} \right], \text{ where } f = 2 \sum_{i=2}^m \binom{p}{i}.$$

3.6. TEST OF SERIAL INDEPENDENCE

The problem of testing for serial independence of a multivariate stationary sequence is now addressed. The test statistic in the serial context is very similar. Consider a stationary sequence Y_1, Y_2, \dots in \mathbb{R}^d , where Y_j is distributed according to the characteristic function g . Let $p \geq 2$ be a fixed integer. Let $m = n - p + 1$ and

$$Z_k = (Y_k, Y_{k+1}, \dots, Y_{k+p-1}) = (Z_k^{(1)}, \dots, Z_k^{(p)}), \quad k = 1, \dots, m, \quad j = 1, \dots, p,$$

where $Z_k^{(j)} = Y_{k+j-1}$. For a given $B \in \mathcal{I}_p$, the process in the serial case is

$$R_{nB,s}(t^{(B)}) = \sqrt{m} \sum_{C \subset B} (-1)^{|B \setminus C|} f_m^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f_m^{(j)}(t^{(j)}),$$

where

$$f_m^{(C)}(t^{(C)}) = \frac{1}{m} \sum_{k=1}^m e^{i \langle t^{(C)}, Z_k^{(C)} \rangle}$$

is the joint empirical characteristic function and

$$f_m^{(j)}(t^{(j)}) = \frac{1}{m} \sum_{k=1}^m e^{i \langle t^{(j)}, Y_{k+j-1} \rangle}, \quad j = 1, \dots, p,$$

are the marginal empirical characteristic functions. Note that all the functions $f_m^{(j)}$ are essentially the same estimate of the unknown characteristic function g . They are not replaced by a single estimate based on all n observations to preserve the representation of the functional (3.5) in terms of doubly-centered matrices. The dependence statistic for the subset B is now defined as the functional

$$T_{nB,s} = \int |R_{nB,s}(t^{(B)})|^2 \prod_{j \in B} dw(t^{(j)}).$$

It can be computed as before using

$$T_{nB,s} = \frac{1}{m} \sum_{k=1}^m \sum_{l=1}^m \prod_{j \in B} A_{kl}^{(j)},$$

where $A_{kl}^{(j)} = a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}$. Two types of weighting measures are considered.

- (1) For *HSIC*, $dw(t^{(j)}) = dG(t^{(j)})$ is a probability measure with characteristic function φ in which case $a_{kl}^{(j)} = \varphi(Z_k^{(j)} - Z_l^{(j)})$. The dependence statistic is denoted $T_{nB,s} = \mathcal{H}_{nB,s}^{2(\alpha)}$.

- (2) For distance covariance, $dw(t^{(j)}) = [C(d, \alpha)|t^{(j)}|_d^{d+\alpha}]^{-1}$ in which case $a_{kl}^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_d^\alpha$. The moment condition $\mathbb{E}|Y_1|_d^\alpha < \infty$ is required for the weak convergence of the functional which is denoted $T_{nB,s} = \mathcal{V}_{nB,s}^{2(\alpha)}$.

Here, the index set of the process $R_{nB,s}$ is the Euclidian space $\mathbb{R}^{d|B|}$.

Theorem 3.7. *Let Y_1, Y_2, \dots be independent and identically distributed with characteristic function g . Then, for any fixed p , the collection of processes $\{R_{nB,s} : A \in \mathcal{B}_p\}$ converges in $C(\mathbb{R}^{d|B|}, \mathbb{C})$ to independent complex Gaussian processes S_B with complex covariance function*

$$\mathbb{E} [S_B(t^{(B)}) \bar{S}_B(s^{(B)})] = \prod_{j \in B} [g(t^{(j)} - s^{(j)}) - g(t^{(j)})g(-s^{(j)})].$$

In the serial context, the subset B and its translate, say $B+k$, lead essentially to the same process. The set \mathcal{I}_p is thus reduced to $\mathcal{B}_p = \{B \in \mathcal{I}_p : 1 \in B\}$ and has now cardinality $2^{p-1} - 1$. In the next theorem, the asymptotic distribution of $T_{nB,s} = \mathcal{H}_{nB,s}^{2(\alpha)}$ is described for any $\alpha \in (0, 2]$.

Theorem 3.8. *Without loss of generality, let $B = \{1, \dots, k\}$. If Y_1, Y_2, \dots are independent and identically distributed, then $T_{nB,s} \rightsquigarrow T_{B,s}$ for each $B \in \mathcal{B}_p$, with $T_{B,s}$ being distributed as*

$$\sum_{i_1=1}^{\infty} \dots \sum_{i_k=1}^{\infty} \lambda_{i_1} \dots \lambda_{i_k} Z_{i_1 \dots i_k}^2,$$

where $Z_{i_1 \dots i_k}$ are independent standard normal variables, and $\lambda_1, \lambda_2, \dots$ are eigenvalues depending only on the distribution of Y_1 and the measure dw .

The adaptive selection method remains the same as in the mutual independence case.

For a given $B \in \mathcal{B}_p$, the p -value \hat{p}_{nB} corresponding to $T_{nB,s}$ can be computed as before using the same Pearson type III approximation. The combined p -value à la Fisher becomes

$$P \left[\chi_f^2 > -2 \sum_{B \in \mathcal{B}_p} \log \hat{p}_{nB} \right], \text{ where } f = 2(2^{p-1} - 1).$$

3.7. SIMULATIONS

For each subset B , the statistic T_{nB} or $T_{nB,s}$, and the corresponding critical value c_B or $c_{B,s}$, are computed. In order to assess the performance of the proposed tests, a dependogram is constructed. Genest and Rémillard (2005) introduced the dependogram which is a graphical tool, with subsets B ordered by size on the horizontal axis, and the corresponding values of T_{nB} or $T_{nB,s}$ represented by vertical bars on the vertical axis. In the mutual independence situation, the

critical value is represented by a dash. Dependence in a subset B is declared when the vertical bar extends beyond the dash. In the serial case, variables selected from subsets B with the same cardinality have the same null distribution. Hence, the same critical value $c_{B,s}$ is used and an horizontal line is drawn for all such subsets.

The global significance level is fixed at $\alpha = .05$. The critical value c_B (or $c_{B,s}$) is the β -quantile, $\beta = (1 - \alpha)^{1/(2^p - p - 1)}$ (or $\beta = (1 - \alpha)^{1/(2^{(p-1)} - 1)}$), of the Pearson type III distribution as an approximation to the permutation distribution of the statistic T_{nB} (or $T_{nB,s}$). Their computation is very fast since approximations do not require any permutation. The test which consists of rejecting the mutual (or serial) independence when at least one vertical bar extends beyond the dash is in fact the global test defined by $\min_{B \in \mathcal{I}_p} \hat{p}_{nB}$. The global test which combines p -values à la Fisher should be preferred. If only subsets B of order up to m are considered in the dependogram, the value of β is modified accordingly to the number of subsets B considered. If this number of subsets is still large and that $1 - \beta$ is very small, then one may set an arbitrary small value of say $1 - \beta = .005$. The dependogram should be seen as an exploratory tool that can be used when Fisher's global test rejects the mutual independence.

A difficulty in interpreting a dependogram is best illustrated by an example. Consider a model with three dependent variables $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$, where $X^{(1)}$ and $X^{(2)}$ are dependent, but the pair $(X^{(1)}, X^{(2)})$ is independent of $X^{(3)}$. One can easily verify that $\mu_{12} \neq 0$, $\mu_{13} \equiv 0$, $\mu_{23} \equiv 0$, and $\mu_{123} \equiv 0$. Therefore, only the test for the subset $\{1, 2\}$ is expected to be significant. The test for the subset $\{1, 2, 3\}$ is not expected to be significant even though the three variables are dependent. Of course, the global test is expected to be significant as it combines the tests for all subsets.

3.7.1. Dependence between three two-dimensional and two three-dimensional vectors

The example considered in this section is taken from Kojadinovic and Holmes (2009) which extends an example in Genest and Rémillard (2005). A sample of size $n = 100$ is generated from the distribution of a 12-dimensional random vector as follows.

- (1) Generate a two-dimensional Gaussian vector $Z^{(1)} = (Z_1^{(1)}, Z_2^{(1)})$ with $Z_i^{(1)} \sim \mathcal{N}(0, 1)$ and $\mathbb{E}(Z_1^{(1)} Z_2^{(1)}) = 0.5$.
- (2) Generate $Z^{(2)}$ and $Z^{(3)}$, two independent copies of $Z^{(1)}$.

- (3) Generate a three-dimensional Gaussian vector $Z^{(4)} = (Z_1^{(4)}, Z_2^{(4)}, Z_3^{(4)})$ with $Z_i^{(4)} \sim \mathcal{N}(0, 1)$ and $\mathbb{E}(Z_i^{(4)} Z_j^{(4)}) = 0.3$, $i \neq j$.
- (4) Generate $Z^{(5)}$, an independent copy of $Z^{(4)}$ such that $(Z^{(4)}, Z^{(5)})$ is independent of $(Z^{(1)}, Z^{(2)}, Z^{(3)})$.
- (5) Define $X^{(1)} = (X_1^{(1)}, X_2^{(1)})$ by $X_i^{(1)} = |Z_i^{(1)}| \text{sign}(Z_1^{(2)} Z_1^{(3)})$, $i = 1, 2$, and set $X^{(2)} = Z^{(2)}$ and $X^{(3)} = Z^{(3)}$.
- (6) Set $X^{(4)} = Z^{(4)}$, and define $X^{(5)} = Z^{(4)} + Z^{(5)}$.

Following Romano and Siegel (1986), the three two-dimensional vectors $(X^{(1)}, X^{(2)}, X^{(3)})$ are pairwise independent, but not jointly independent. This vector is independent of $(X^{(4)}, X^{(5)})$ in which the two three-dimensional vectors $X^{(4)}$ and $X^{(5)}$ are dependent. Figure 3.1 is the dependogram based on one simulated sample. The test statistics computed are the distance covariance and the adaptive *HSIC*, both of index 1. The dependence among $X^{(4)}$ and $X^{(5)}$, represented by the subset $\{4, 5\}$, is significant. Moreover, the third order dependence between $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$, represented by the subset $\{1, 2, 3\}$, is also significant. The two global tests have p -values equal to zero. For both tests, the computations were very fast and took less than 0.1 s on an Intel(R) Core(TM) i7 with a CPU of 3.20 GHz running with Linux.

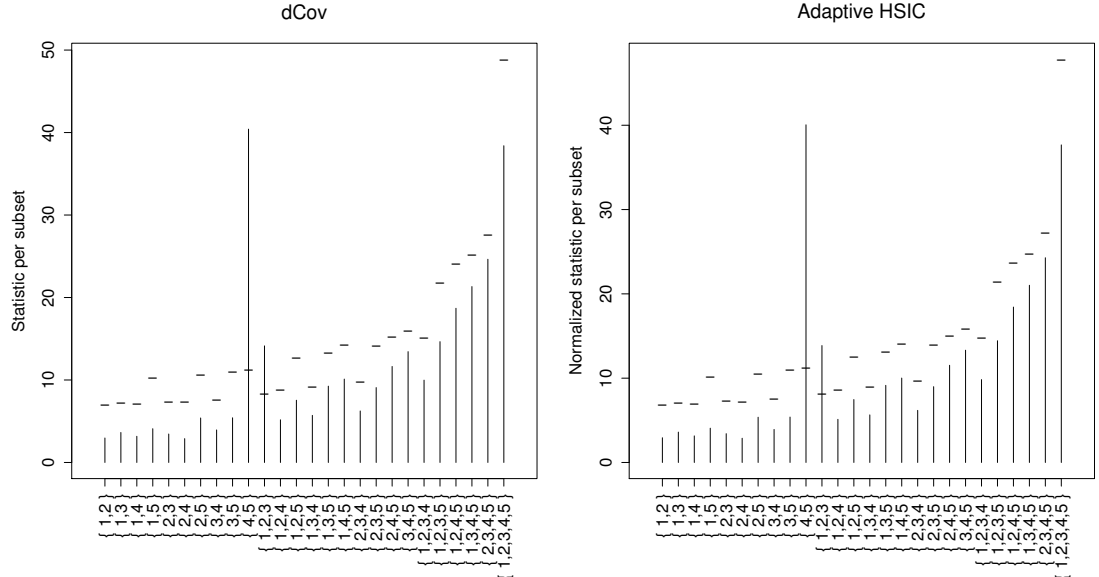


FIGURE 3.1. Dependogram between three two-dimensional and two three-dimensional vectors as in Kojadinovic and Holmes (2009). The sample size is $n = 100$. In the right panel, the *HSIC* test statistics are normalized by $\prod_{j \in A} \hat{\beta}_j$ as in Equation (3.8).

Type I error rates or powers reported in Table 3.1 were estimated from 1000 tests at nominal level 5%. *HSIC* tests and distance covariance tests reach the nominal level for all subsets B in which all elements are mutually independent. The subsets $\{4, 5\}$, $\{1, 2, 3\}$, and the global test detect the dependence with a power of 1.

TABLE 3.1. Empirical error rates and powers of *HSIC* and distance covariance tests of nominal level 5% between three two-dimensional and two three-dimensional vectors as in Kojadinovic and Holmes (2009). A subset is represented by the letter B . The sample size is $n = 100$. The number of replications is 1000.

B	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{1,5\}$	$\{2,3\}$	$\{2,4\}$	$\{2,5\}$
\mathcal{V}_{nB}^2	0.045	0.051	0.047	0.047	0.049	0.044	0.042
\mathcal{H}_{nB}^2	0.045	0.051	0.047	0.047	0.049	0.044	0.042
B	$\{3,4\}$	$\{3,5\}$	$\{4,5\}$	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,2,5\}$	$\{1,3,4\}$
\mathcal{V}_{nB}^2	0.055	0.044	1	1	0.056	0.055	0.049
\mathcal{H}_{nB}^2	0.055	0.044	1	1	0.057	0.055	0.049
B	$\{1,3,5\}$	$\{1,4,5\}$	$\{2,3,4\}$	$\{2,3,5\}$	$\{2,4,5\}$	$\{3,4,5\}$	$\{1,2,3,4\}$
\mathcal{V}_{nB}^2	0.043	0.127	0.047	0.059	0.164	0.156	0.029
\mathcal{H}_{nB}^2	0.042	0.130	0.047	0.059	0.165	0.158	0.029
B	$\{1,2,3,5\}$	$\{1,2,4,5\}$	$\{1,3,4,5\}$	$\{2,3,4,5\}$	$\{1,2,3,4,5\}$	Global test	
\mathcal{V}_{nB}^2	0.029	0.310	0.314	0.323	0.525	1	
\mathcal{H}_{nB}^2	0.029	0.312	0.315	0.325	0.524	1	

3.7.2. Empirical power for data from a Student copula

Another example from Kojadinovic and Holmes (2009) is slightly modified. Data of sample size $n = 100$ are simulated from a Student copula of dimension 6 with 2 degrees of freedom and correlation matrix constructed as follows. A partial correlation matrix is first constructed of the form

$$P = \begin{pmatrix} (1 - \rho_w)I_2 + \rho_w J_2 J_2' & \rho_b J_2 J_2' & \rho_b J_2 J_2' \\ \rho_b J_2 J_2' & (1 - \rho_w)I_2 + \rho_w J_2 J_2' & \rho_b J_2 J_2' \\ \rho_b J_2 J_2' & \rho_b J_2 J_2' & (1 - \rho_w)I_2 + \rho_w J_2 J_2' \end{pmatrix},$$

where I_2 is the identity matrix of dimension 2, and J_2 is the vector of ones of dimension 2. Notations w and b stand for within and between, respectively. Then, the one-to-one transformation of Joe (2006) between partial correlations and simple correlations is applied to P to yield a correlation matrix R . The resulting matrix R is positive definite for all ρ_w and ρ_b in the interval $(-1, 1)$, although P may not be positive definite. Probability transforms are finally applied so that all 6 variables become uniformly distributed on the interval $(0, 1)$. The resulting

vector is partitioned into three two-dimensional vectors. The values of ρ_w are: 0 and 0.5. For each value of ρ_w , we consider the value of ρ_b : 0, 0.05, 0.1, 0.15 and 0.2.

Two distance covariance tests of index 1/2 and 1 were compared to two *HSIC* tests of index 1/2: $\mathcal{H}_n^{2(1/2)}$ with very small scale parameters $\beta_j = 10^{-4}/\text{med}_{k<l}|Z_k^{(j)} - Z_l^{(j)}|$ and $\mathcal{H}_{n,\text{adap}}^{2(1/2)}$ with adaptive selection. The notations $\mathcal{V}_n^{2(1/2)}$, \mathcal{V}_n^2 , $\mathcal{H}_n^{2(1/2)}$ and $\mathcal{H}_{n,\text{adap}}^{2(1/2)}$ represent the global tests obtained by combining p -values à la Fisher. The test of Kojadinovic and Holmes (2009) was not considered since it performed poorly in Chapter 2 for Gaussian data in the case $p = 2$ as compared to the test of distance covariance. Figure 3.2 shows that the best distance covariance test has index 1/2 and it is matched by $\mathcal{H}_n^{2(1/2)}$. However, the test $\mathcal{H}_{n,\text{adap}}^{2(1/2)}$ with adaptive selection has the best power, especially for small values of ρ_b .

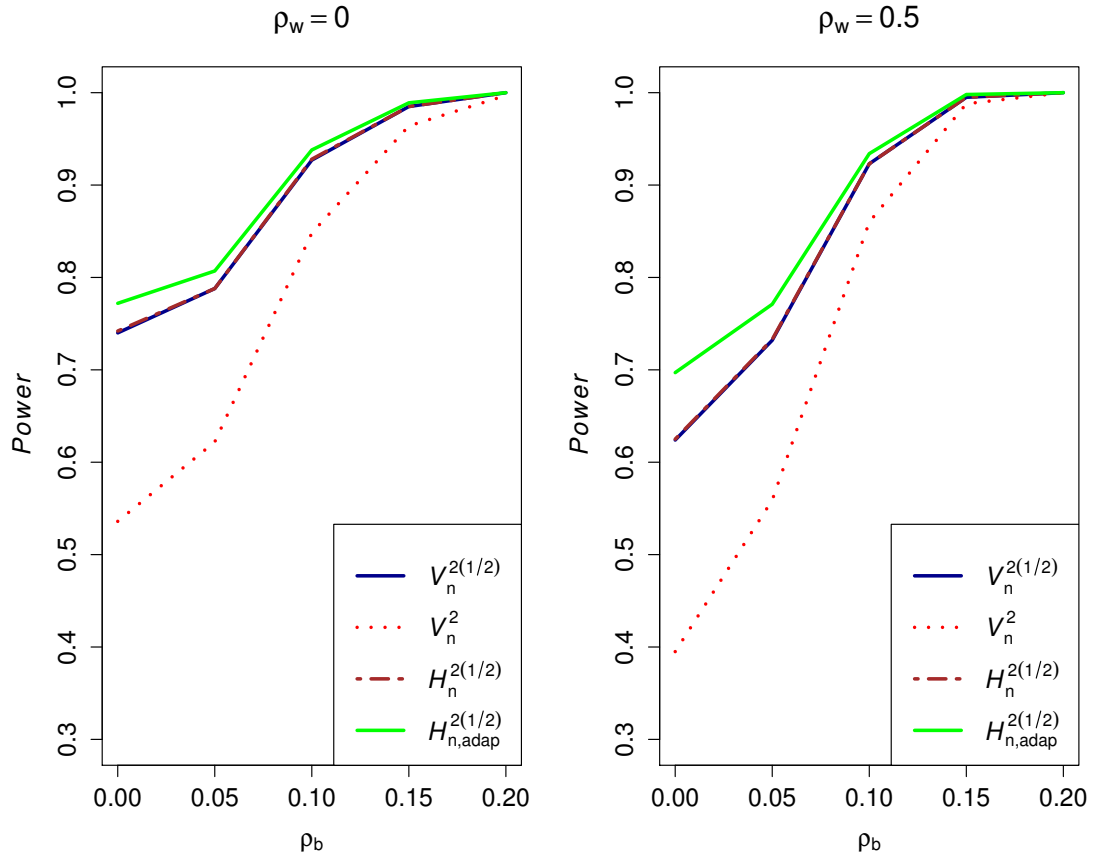


FIGURE 3.2. Power functions simulated from 1000 tests at 5% significance level. Two distance covariance tests of index 1/2 and 1 are compared to two *HSIC* tests of index 1/2. Data follow a Student copula of dimension 6 with two degrees of freedom. The sample size is $n = 100$.

For the Student copula, the three vectors are dependent for $\rho_b = 0$. Hence, empirical type I error rates in Table 3.2 were verified for three independent Student copulas. They are estimated from 1000 tests at nominal level 5%. Data are simulated for the two cases of intra correlation $\rho_w = 0$ and $\rho_w = 0.5$. Analysis of empirical type I error rates is made for all the possible subsets and for the global tests. For a subset B , two distance covariance tests $\mathcal{V}_{nB}^{2(1/2)}$ and \mathcal{V}_{nB}^2 of index 1/2 and 1, respectively, and two *HSIC* tests of index 1/2: $\mathcal{H}_{nB}^{2(1/2)}$ with very small scale parameters $\beta_j = 10^{-4}/\text{med}_{k < l} |Z_k^{(j)} - Z_l^{(j)}|$ and $\mathcal{H}_{nB, \text{adap}}^{2(1/2)}$ with adaptive selection are considered. The corresponding global tests as in Figure 3.2 are also considered. Simulations in Table 3.2 show that all the tests reach the nominal level of 5% for each subset and for the global test.

TABLE 3.2. Empirical type I error rates simulated from 1000 tests at 5% significance level. Two distance covariance tests of index 1/2 and 1 and two *HSIC* tests of index 1/2 are considered. Simulations are based on three independent bivariate Student copulas.

$\rho_w = 0$						
B	{1,2}	{1,3}	{2,3}	{1,2,3}	Global test	
$\mathcal{V}_{nB}^{2(1/2)}$	0.054	0.046	0.044	0.040	$\mathcal{V}_n^{2(1/2)}$	0.048
\mathcal{V}_{nB}^2	0.050	0.043	0.053	0.042	\mathcal{V}_n^2	0.047
$\mathcal{H}_{nB}^{2(1/2)}$	0.054	0.045	0.044	0.040	$\mathcal{H}_n^{2(1/2)}$	0.048
$\mathcal{H}_{nB, \text{adap}}^{2(1/2)}$	0.050	0.045	0.042	0.041	$\mathcal{H}_{n, \text{adap}}^{2(1/2)}$	0.050
$\rho_w = 0.5$						
B	{1,2}	{1,3}	{2,3}	{1,2,3}	Global test	
$\mathcal{V}_{nB}^{2(1/2)}$	0.050	0.047	0.043	0.056	$\mathcal{V}_n^{2(1/2)}$	0.042
\mathcal{V}_{nB}^2	0.043	0.055	0.040	0.055	\mathcal{V}_n^2	0.044
$\mathcal{H}_{nB}^{2(1/2)}$	0.050	0.046	0.043	0.056	$\mathcal{H}_n^{2(1/2)}$	0.042
$\mathcal{H}_{nB, \text{adap}}^{2(1/2)}$	0.048	0.044	0.041	0.056	$\mathcal{H}_{n, \text{adap}}^{2(1/2)}$	0.040

3.7.3. Serial dependence of a binary sequence of zeros and ones

This example is taken from Beran et al. (2007). An independent sequence V_1, \dots, V_n of length $n = 100$ is simulated so that variables V_i take values 0 and 1 with probabilities 0.2 and 0.8, respectively. A sequence, dependent at lag 3, is created by the product $Z_i = V_i V_{i+3}$, $i = 1, \dots, n-3$. The distance covariance test and the adaptive *HSIC* test, both of index 1 and with $p = 4$, were used in the dependograms in Figure 3.3. The value $p = 4$ is the minimal value required to detect a dependence at lag 3. The dependograms took less than 0.4 s to compute. In the top row, there is no evidence of any serial dependence of the independent sequence V_i . However, in the bottom row, a significant serial dependence at lag 3 is detected in the sequence Z_i through the subset $\{1, 4\}$. The computation times

for tests of Kojadinovic and Yan (2011) not reported here were approximately 20 times longer.

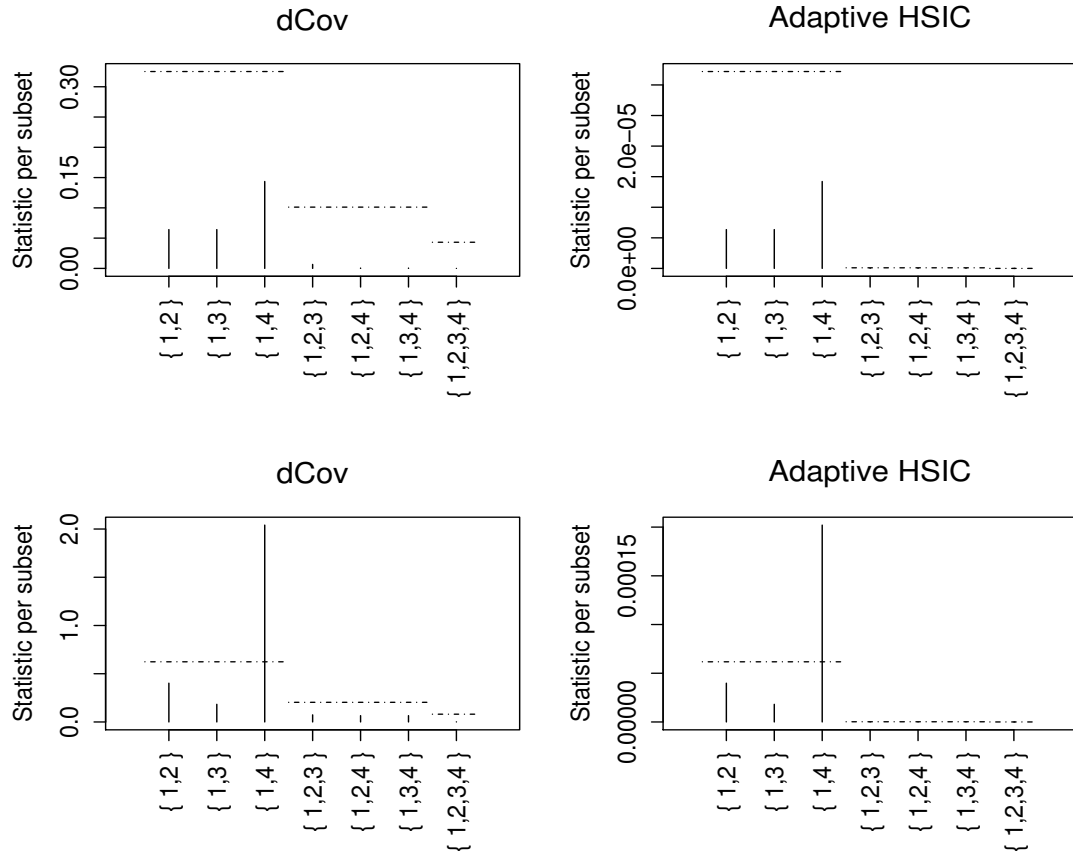


FIGURE 3.3. The top row exhibits dependograms for the independent sequence V_i . The bottom row contains dependograms for the sequence Z_i which is dependent at lag 3. The tests used are the distance covariance in the left panel and the adaptive *HSIC* in the right panel, both of index 1.

Empirical type I error rates for the independent sequence V_i and the empirical powers for the lag 3 dependent sequence Z_i are reported in Table 3.3. They are estimated from 1000 tests at nominal level 5%. For the independent sequence V_i , *HSIC* test and distance covariance tests, both of index 1, reach the nominal level for all the possible subsets and for the global test. The empirical powers of the global tests for the dependent sequence is about 0.944 for both *HSIC* and distance covariance tests.

TABLE 3.3. Empirical type I error rates for the independent sequence V_i and the empirical powers for the lag 3 dependent sequence Z_i . The tests used are the distance covariance and the adaptive *HSIC*, both of index 1. The length of the sequences is $n = 100$ each. The number of replications is 1000.

Independent sequence V_i				
B	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{1,2,3\}$
\mathcal{V}_{nB}^2	0.051	0.052	0.056	0.052
\mathcal{H}_{nB}^2	0.051	0.052	0.056	0.052
B	$\{1,2,4\}$	$\{1,3,4\}$	$\{1,2,3,4\}$	Global test
\mathcal{V}_{nB}^2	0.041	0.050	0.037	0.053
\mathcal{H}_{nB}^2	0.041	0.050	0.037	0.053
Dependent sequence Z_i				
B	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{1,2,3\}$
\mathcal{V}_{nB}^2	0.096	0.107	0.998	0.199
\mathcal{H}_{nB}^2	0.096	0.107	0.998	0.199
B	$\{1,2,4\}$	$\{1,3,4\}$	$\{1,2,3,4\}$	Global test
\mathcal{V}_{nB}^2	0.033	0.020	0.099	0.944
\mathcal{H}_{nB}^2	0.033	0.020	0.099	0.944

3.8. APPLICATIONS

3.8.1. Testing mutual independence between air temperature, soil temperature, humidity, wind and evaporation

Data for this section are from Rencher (1995, p. 294). Table 3.4 describes the variables of the data with 46 observations. The R package MVN (Korkmaz et al., 2014) has the function `hzTest` to perform the test for multivariate normality of Henze and Zirkler (1990). This test applied to the joint distribution of all 11 variables rejected a multivariate Gaussian model with a p -value of 0. Now, five groups of variables are considered: air temperature (y_1, y_2, y_3), soil temperature (y_4, y_5, y_6), relative humidity (y_7, y_8, y_9), wind y_{10} and evaporation y_{11} . The Gaussian likelihood ratio test found a significant mutual dependence between the five groups. However, this test should not be relied on since it was found that data are not jointly Gaussian and it is now well known that the Gaussian likelihood ratio test is very non robust (Tyler, 1983; Bilodeau and Brenner, 1999).

Mutual independence between these five groups is now tested with the distance covariance and the adaptive *HSIC* tests, both of index 1. Figure 3.4 contains dependograms only for subsets B of order up to 3. It reveals that air temperature, soil temperature, relative humidity and evaporation are pairwise dependent.

TABLE 3.4. Variables related to air temperature, soil temperature, humidity, wind and evaporation.

Variables	Labels
y_1	maximum daily air temperature
y_2	minimum daily air temperature
y_3	integrated area under daily air temperature curve, i.e. a measure of average air temperature
y_4	maximum daily soil temperature
y_5	minimum daily soil temperature
y_6	integrated area under soil temperature
y_7	maximum daily relative humidity
y_8	minimum daily relative humidity
y_9	integrated area under daily humidity curve
y_{10}	total wind, measured in miles per day
y_{11}	evaporation

However, wind does not exhibit any dependence of order 2 or 3 with any of the other 4 groups.

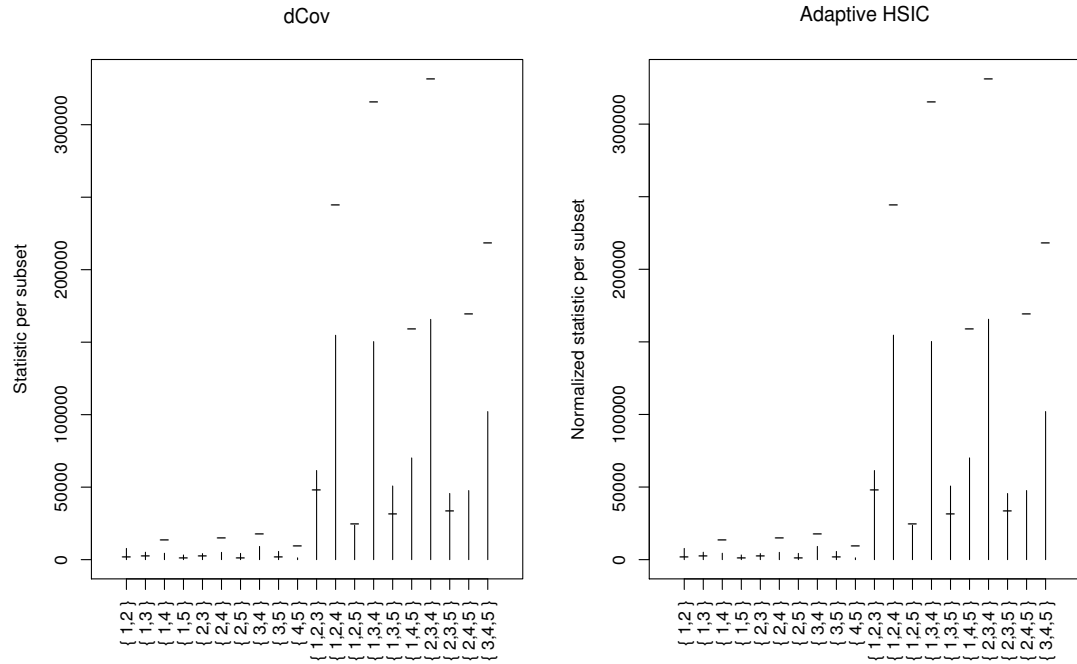


FIGURE 3.4. Dependograms for mutual independence of air temperature, soil temperature, relative humidity, wind and evaporation. In the right panel, the adaptive $HSIC$ test statistics are normalized by $\prod_{j \in A} \hat{\beta}_j$ as in Equation (3.8).

3.8.2. Testing serial independence for financial data

Tests of serial independence of the three dimensional sequence formed by the increasing rate of daily series of S&P/TSX composite (TSX), S&P500 and DOW JONES indices are considered. The values of indices are taken at closure. The series of length 534 range from January 2, 2014 to March 2, 2016. Note that five index values are observed weekly since the stock exchanges are not opened on weekends. The top row of Figure 3.5 shows that the financial series considered are not stationary. It is more appropriate to consider the increase rates of the series. In the bottom row of Figure 3.5, one may see that the increasing rates can be considered as stationary.

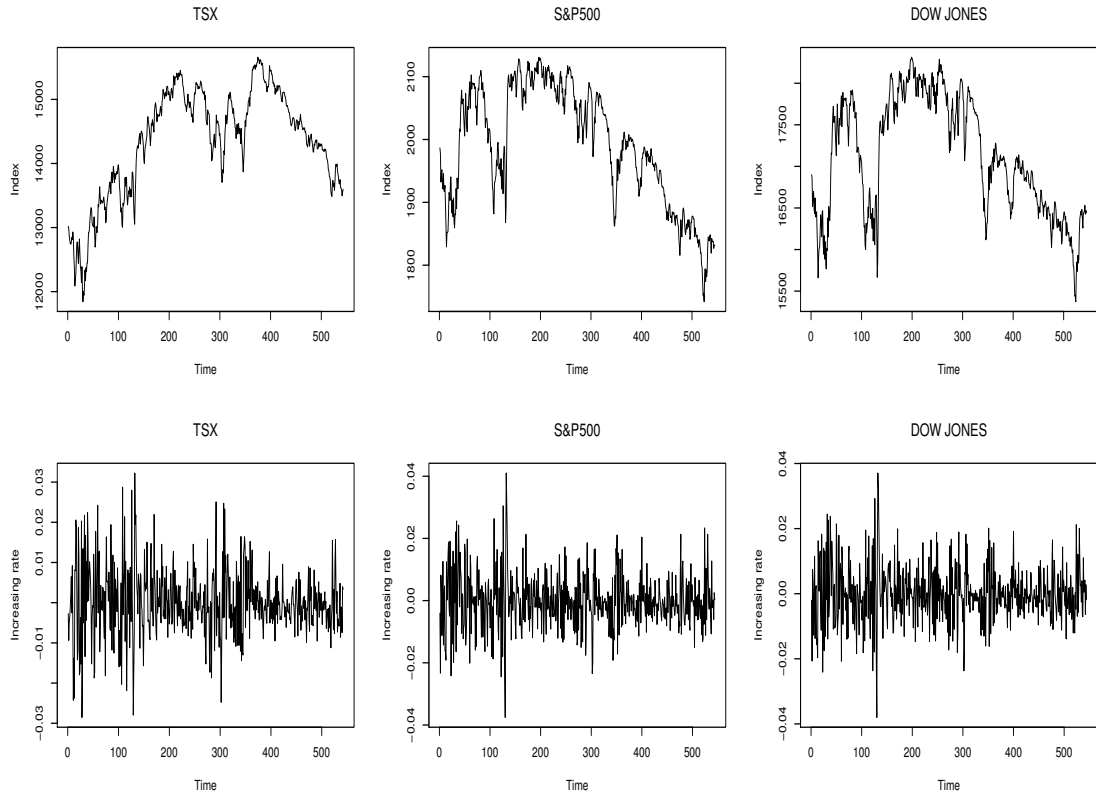


FIGURE 3.5. Evolution of three daily financial series and their increasing rate taken at their closure value. The period of observation ranges from January 2, 2014 to March 2, 2016.

The adaptive *HSIC* test and the distance covariance test, both of index 1, are conducted on the 3 series jointly. The maximum lag is $p = 100$ which covers about one quarter. In Figure 3.6, the adaptive *HSIC* reveals dependencies at small lags of 1, 2, 3, and 4. Of these four dependencies, only the dependence at lag 4 is declared by the distance covariance. This may be the result of the

superiority of the adaptive *HSIC* over distance covariance in terms of powers. The adaptive *HSIC* test also reveals dependencies at lags 8 to 11, 15, 16, and 78. The two global tests have p -values equal to zero.

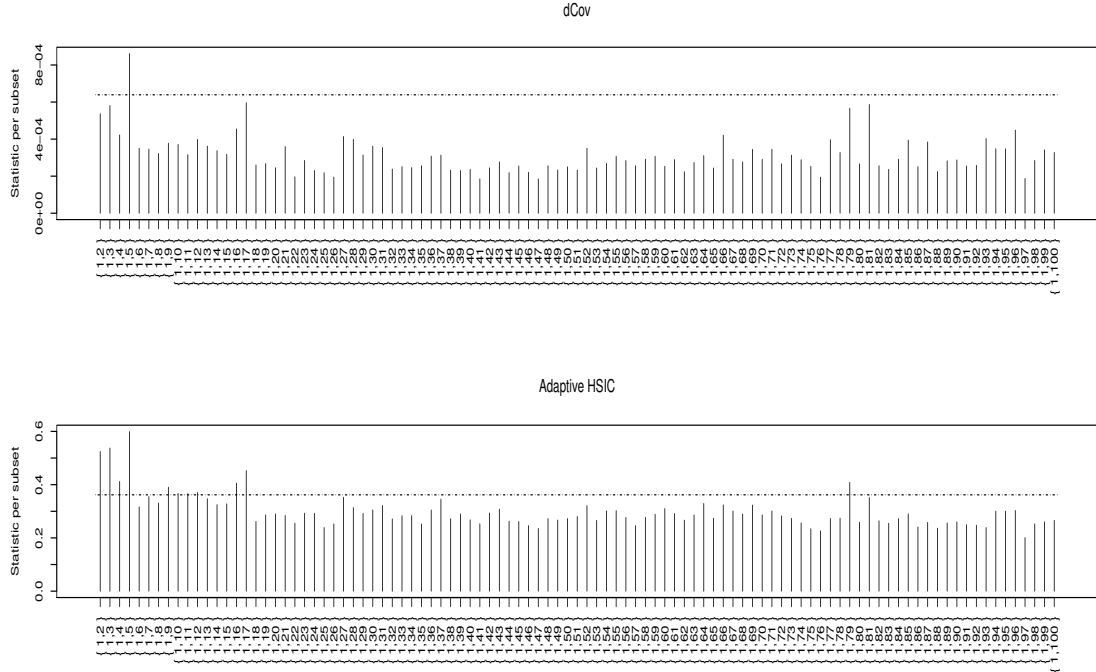


FIGURE 3.6. Dependograms of the serial dependence tests for the multivariate financial sequence of dimension 3. The multivariate sequence is formed by the increasing rates of the S&P/TSX composite, the S&P500 and the DOW JONES indices. The maximum lag is $p = 100$. Two tests are considered: distance covariance in the top and adaptive *HSIC* in the bottom, both of index 1.

3.9. CONCLUSION

Generalizations of the distance covariance test and the *HSIC* test were done successfully. For both mutual and serial independence tests, the dependence statistics related to distance covariance and *HSIC* were defined using the Möbius transformation. Simple and explicit expressions for dependence statistics were derived in the generic form (3.5) as a sum over all elements of a componentwise product of doubly-centered matrices. The exact first three moments of the permutation distribution of any test of this form were derived and used in a Pearson type III approximation to compute p -values very efficiently without resorting to any resampling method. The method of combining p -values à la Fisher was put forward for testing the global hypotheses of mutual and serial independence. The

adaptive method for the selection of scale parameters of *HSIC* tests applied to data generated from a Student copula improved the power of the *HSIC* test compared to the distance covariance test.

The tests proposed are only invariant to orthogonal transformations but they have a very low computational cost. Another test of Beran et al. (2007) of the Kolmogorov-Smirnov type are invariant to the general linear group of transformations. However, the test statistics are obtained by a very costly numerical optimization and p -values are estimated by the bootstrap. They would require several hours of computations for the meteorological data set of Section 3.8.1.

APPENDIX A: PROOFS

PROOF OF THEOREM 3.1. Upon using the representation (3.3) of the process,

$$\begin{aligned}
& \int |R_{nB}(t^{(B)})|^2 \prod_{j \in B} dG^{(j)}(t^{(j)}) \\
&= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \prod_{j \in B} \int \left[\exp(i\langle t^{(j)}, Z_k^{(j)} - Z_l^{(j)} \rangle) - \frac{1}{n} \sum_{v=1}^n \exp(i\langle t^{(j)}, Z_k^{(j)} - Z_v^{(j)} \rangle) \right. \\
&\quad \left. - \frac{1}{n} \sum_{u=1}^n \exp(i\langle t^{(j)}, Z_u^{(j)} - Z_l^{(j)} \rangle) + \frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \exp(i\langle t^{(j)}, Z_u^{(j)} - Z_v^{(j)} \rangle) \right] dG^{(j)}(t^{(j)}) \\
&= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \prod_{j \in A} [a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}],
\end{aligned} \tag{3.14}$$

where $a_{kl}^{(j)} = \varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)})$ and $\varphi^{(j)}$ is the characteristic function of $G^{(j)}$. \square

PROOF OF THEOREM 3.2. Because of double-centering, the expression between brackets in (3.14) is unchanged if one is subtracted from all four exponential functions. Then, it suffices in the proof of Theorem 3.1, for the weight function defining $\mathcal{V}_n^{2(\alpha)}$, to evaluate

$$\int \frac{\exp(i\langle t^{(j)}, Z_k^{(j)} - Z_l^{(j)} \rangle) - 1}{C(d_j, \alpha) |t^{(j)}|_{d_j}^{d_j + \alpha}} dt^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha$$

with Lemma 1 of Székely et al. (2007, p. 2771). \square

The representation (3.3) of the process was unexploited in Székely et al. (2007). It simplifies greatly their derivations.

PROOF OF EQUATION (3.8). The result follows using the invariance by translation, $a_{kl}^{(j)} \mapsto a_{kl}^{(j)} - 1$, and the following limit,

$$\lim_{\beta_j \rightarrow 0} \frac{e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha} - 1}{\beta_j^\alpha} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha.$$

Let $d_B = \sum_{j \in B} d_j$. Define the metric (Whitt, 1970)

$$\rho(x, y) = \sum_{s=1}^{\infty} 2^{-s} \frac{\rho_s(x, y)}{1 + \rho_s(x, y)},$$

where

$$\rho_s(x, y) = \sup_{|t^{(B)}|_{d_B} \leq s} |x(t^{(B)}) - y(t^{(B)})|,$$

on the linear complete metric space of continuous functions $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$. The Borel σ -field in $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$ is generated by the coordinate projections, *i.e.* it is the smallest σ -field with respect to which all coordinate projections are measurable. Weak convergence of random variables in $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$ is equivalent to weak convergence on any compact subset (Kallenberg, 2002, Proposition 16.6). Moreover, weak convergence of a sequence on a compact subset is equivalent to finite dimensional weak convergence and tightness of that sequence.

PROOF OF THEOREM 3.3. The process (3.3) is replaced by the closely related process

$$\check{R}_{nB}(t^{(B)}) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \prod_{j \in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)} \rangle} - f^{(j)}(t^{(j)}) \right].$$

in which marginal characteristic functions are not estimated. The process \check{R}_{nB} is a sum of independent and identically distributed random variables. Bilodeau and Lafaye de Micheaux (2005, Theorem 2.1) proved that the collection of processes \check{R}_{nB} converges as stated in Theorem 3.3 under the weak condition (3.10) in Csörgő (1981). The independence of the asymptotic processes for $B \neq C$ is verified

$$\begin{aligned} \mathbb{E} \left[R_B(t^{(B)}) \bar{R}_C(s^{(C)}) \right] &= \mathbb{E} \left\{ \prod_{j \in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)} \rangle} - f^{(j)}(t^{(j)}) \right] \right. \\ &\quad \cdot \left. \prod_{j \in C} \left[e^{i\langle s^{(j)}, -Z_k^{(j)} \rangle} - f^{(j)}(-s^{(j)}) \right] \right\} \\ &= 0, \end{aligned}$$

because there is an index j in B , but not in C , or the converse, for which the corresponding term has expectation zero. Then, it suffices to show $\rho_s(R_{nB}, \check{R}_{nB}) \xrightarrow{P} 0$, for all $s \geq 1$. Following Ghoudi et al. (2001, p. 212), observe that

$$|R_{nB}(t^{(B)}) - \check{R}_{nB}(t^{(B)})| \leq \sum_{C \subset B, C \neq \emptyset} \prod_{j \in C} |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})| |\check{R}_{n, B \setminus C}(t^{(B \setminus C)})|, \quad (3.15)$$

where the sum has only a finite number of terms. Using the Glivenko-Cantelli convergence in Csörgő (1981, Theorem 2.1) and the fact that the processes $\check{R}_{n,B \setminus C}$ are tight, it follows that, for any $s \geq 1$, $\rho_s(R_{nB}, \check{R}_{nB}) \xrightarrow{P} 0$. \square

PROOF OF THEOREM 3.4. The proof consists in showing the following two results: $\check{T}_{nA} \rightsquigarrow T_B$ and $|\check{T}_{nB} - T_{nB}| \xrightarrow{P} 0$, where

$$\check{T}_{nB} = \int |\check{R}_{nB}(t^{(B)})|^2 dw_B.$$

Using a slight generalization of Kellermeier (1980, Theorem 3.3), it suffices for the first result to show the following uniform integrability condition:

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \int_{\{|t^{(j)}| > N, \forall j \in B\}} \mathbb{E} |\check{R}_{nB}(t^{(B)})|^2 \prod_{j \in B} dw^{(j)}(t^{(j)}) = 0.$$

It can be easily verified that $\mathbb{E} |\check{R}_{nB}(t^{(B)})|^2 = \prod_{j \in B} [1 - |f^{(j)}(t^{(j)})|^2]$ does not depend on n . Now, for each $j \in B$,

$$\int 1 - |f^{(j)}(t^{(j)})|^2 dw^{(j)}(t^{(j)}) < \infty. \quad (3.16)$$

Indeed, for $\mathcal{H}_{nB}^{2(\alpha)}$, $dw^{(j)}(t^{(j)})$ is a probability measure and the integrand is bounded. For the second result, it suffices by the Cauchy-Schwarz inequality to show

$$\int |R_{nB}(t^{(B)}) - \check{R}_{nB}(t^{(B)})|^2 \prod_{j \in B} dw^{(j)}(t^{(j)}) \xrightarrow{P} 0.$$

From (3.15), since all processes $\check{R}_{n,B \setminus C}$ are tight, it suffices to show that for all $j \in B$,

$$\int |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})|^2 dw^{(j)}(t^{(j)}) \xrightarrow{P} 0.$$

Again, from Kellermeier (1980, Theorem 3.3), it suffices to show the uniform integrability condition:

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \int_{|t^{(j)}| > N} \mathbb{E} |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})|^2 dw^{(j)}(t^{(j)}) = 0. \quad (3.17)$$

It may be seen that $\mathbb{E} |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})|^2 = (1 - |f^{(j)}(t^{(j)})|^2)/n$. Hence, (3.17) follows from (3.16). Finally, the representation (3.10) for T_B follows from the usual Karhunen-Loève expansion. The product structure of eigenvalues is inherited from the product structure of the covariance (3.9) and of the product measure dw_B . \square

A symbol like \sum'_{ijk} will represent a sum over indices i, j, k all different.

PROOF OF THEOREM 3.5.

$$(n!)^p \mathbb{E}_P(T) = \sum_{kl} \prod_{a=1}^p \sum_{\sigma_a} A_{kl}^{(a)}(\sigma_a)$$

$$= \sum_k \prod_{a=1}^p \sum_{\sigma_a} A_{kk}^{(a)}(\sigma_a) + \sum'_{kl} \prod_{a=1}^p \sum_{\sigma_a} A_{kl}^{(a)}(\sigma_a).$$

A direct evaluation gives

$$\begin{aligned} \sum_{\sigma_a} A_{kk}^{(a)}(\sigma_a) &= (n-1)! \sum_k A_{kk}^{(a)} \\ &= (n-1)! T^{(a)}, \\ \sum_{\sigma_a} A_{kl}^{(a)}(\sigma_a) &= (n-2)! \sum'_{kl} A_{kl}^{(a)} \\ &= (n-2)! \left[\sum_{kl} A_{kl}^{(a)} - \sum_k A_{kk}^{(a)} \right] \\ &= (n-2)! \left[-T^{(a)} \right], \quad k \neq l. \end{aligned}$$

It follows

$$(n!)^{p-1} \mathbb{E}_P T = [(n-1)!]^{p-1} \prod_{a=1}^p T^{(a)} + [(n-2)!]^{p-1} \prod_{a=1}^p \left[-T^{(a)} \right].$$

□

Two lemmas are needed for the formula of the second moment.

Lemma 3.1 (Sums for one or two distinct indices). *The following expressions hold*

$$\begin{aligned} \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 &= (n-1)! S_2^{(a)}, \\ \sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) &= (n-2)! \left[(T^{(a)})^2 - S_2^{(a)} \right], \\ \sum_{\sigma_a} [A_{ij}^{(a)}(\sigma_a)]^2 &= (n-2)! \left[T_2^{(a)} - S_2^{(a)} \right], \\ \sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) &= (n-2)! \left[-S_2^{(a)} \right]. \end{aligned}$$

Lemma 3.2 (Sums for three or four distinct indices). *The following expressions hold*

$$\begin{aligned} \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) &= (n-3)! \left[-T_2^{(a)} + 2S_2^{(a)} \right], \\ \sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jk}^{(a)}(\sigma_a) &= (n-3)! \left[-(T^{(a)})^2 + 2S_2^{(a)} \right], \\ \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) &= (n-4)! \left[(T^{(a)})^2 + 2T_2^{(a)} - 6S_2^{(a)} \right]. \end{aligned} \quad (3.18)$$

PROOFS OF LEMMAS 3.1 AND 3.2. Equation (3.18) is proved when i, j, k, l are four distinct indices. The sums over all permutations are

$$\begin{aligned}
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) &= (n-4)! \sum'_{ijkl} A_{ij}^{(a)} A_{kl}^{(a)} \\
&= (n-4)! \left[\sum'_{ij} A_{ij}^{(a)} \sum_{\substack{k \neq i \\ k \neq j}} (-A_{ki}^{(a)} - A_{kj}^{(a)} - A_{kk}^{(a)}) \right] \\
&= -(n-4)! \left[\sum'_{ij} A_{ij}^{(a)} (T^{(a)} - 2A_{ii}^{(a)} - 2A_{jj}^{(a)} - 2A_{ij}^{(a)}) \right] \\
&= (n-4)! \left[-T^{(a)} \sum'_{ij} A_{ij}^{(a)} + 4 \sum'_{ij} A_{ii}^{(a)} A_{ij}^{(a)} + 2 \sum'_{ij} (A_{ij}^{(a)})^2 \right] \\
&= (n-4)! \left[(T^{(a)})^2 - 4S_2^{(a)} + 2(T_2^{(a)} - S_2^{(a)}) \right] \\
&= (n-4)! \left[(T^{(a)})^2 + 2T_2^{(a)} - 6S_2^{(a)} \right].
\end{aligned}$$

The other sums over all permutations in Lemmas 3.1 and 3.2 are calculated similarly. \square

PROOF OF THEOREM 3.6.

$$\begin{aligned}
(n!)^p \mathbb{E}_P(T^2) &= \sum_{ijkl} \prod_{a=1}^p \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) \\
&= \sum_i \prod_{a=1}^p \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 \\
&\quad + \sum'_{ij} \prod_{a=1}^p \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) + 2[A_{ij}^{(a)}(\sigma_a)]^2 + 4A_{ii}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a)] \\
&\quad + \sum'_{ijk} \prod_{a=1}^p \sum_{\sigma_a} [4A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) + 2A_{ii}^{(a)}(\sigma_a) A_{jk}^{(2)}(\sigma_a)] \\
&\quad + \sum'_{ijkl} \prod_{a=1}^p \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a). \tag{3.19}
\end{aligned}$$

Sums over permutations in Lemmas 3.1 and 3.2 do not depend on the values of indices, but only on the number of different indices selected. From equation (3.19) and Lemmas 3.1 and 3.2, the second moment in Theorem 3.6 follows. \square

PROOF OF THEOREM 3.7. Consider the processes

$$\check{R}_{nB,s}(t^{(B)}) = \frac{1}{\sqrt{m}} \sum_{k=1}^m \prod_{j \in B} \left[e^{i \langle t^{(j)}, Z_k^{(j)} \rangle} - g(t^{(j)}) \right], \quad B \in \mathcal{B}_p.$$

Finite dimensional weak convergence of the processes is proved. Because of overlapping of Y 's in consecutive Z_k 's, the Z_k 's form an $(p-1)$ -dependent sequence, see Ferguson (1996, p. 69). Thus, the central limit theorem for such dependent sequences establishes that $\check{R}_{nB,s}(t^{(B)})$ and $\check{R}_{nC,s}(s^{(C)})$ are asymptotically and jointly normal with asymptotic covariance $\sigma_{0,0} + 2\sigma_{0,1} + \dots + 2\sigma_{0,p-1}$, where

$$\sigma_{0,u} = \mathbb{E} \left\{ \prod_{j \in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)} \rangle} - g(t^{(j)}) \right] \prod_{j \in C} \left[e^{i\langle s^{(j)}, -Z_{k+u}^{(j)} \rangle} - g(-s^{(j)}) \right] \right\}.$$

All of the above expectations are null unless $B = C$ (both in \mathcal{B}_p) and $u = 0$. Next, to establish weak convergence of the process on any compact, assume without loss of generality that m is a multiple of p , say $m = rp$. This amounts to neglecting at most $p-1$ terms in the sequence. Rewrite the sequence Z_1, Z_2, \dots as an array with p rows, each consisting of r independent and identically distributed vectors,

$$\begin{array}{cccc} Z_1 & Z_{1+p} & \cdots & Z_{1+(r-1)p} \\ Z_2 & Z_{2+p} & \cdots & Z_{2+(r-1)p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_p & Z_{p+p} & \cdots & Z_{p+(r-1)p} \end{array}$$

Then, the expression

$$\check{R}_{nB,s}(t^{(B)}) = \frac{1}{\sqrt{p}} \sum_{h=1}^p \sum_{C \subset B} (-1)^{|B \setminus C|} \prod_{j \in B \setminus C} g(t^{(j)}) \cdot \frac{1}{\sqrt{r}} \sum_{i=0}^{r-1} \left[e^{i\langle t^{(C)}, Z_{pi+h}^{(C)} \rangle} - \prod_{j \in C} g(t^{(j)}) \right]$$

establishes weak convergence since for each pair (h, C) in finite number, the last sum over i is an empirical characteristic function process over a compact. Finally, $R_{nB,s}$ and $\check{R}_{nB,s}$ are equivalent processes follows from the inequality

$$|R_{nB,s}(t^{(B)}) - \check{R}_{nB,s}(t^{(B)})| \leq \sum_{C \subset B, C \neq \emptyset} \prod_{j \in C} |f_n^{(j)}(t^{(j)}) - g(t^{(j)})| |\check{R}_{nB \setminus C,s}(t^{(B \setminus C)})|,$$

and the same arguments following Equation (3.15). \square

PROOF OF THEOREM 3.8. The proof is the same as for Theorem 3.4. \square

APPENDIX B: THIRD MOMENT

The third moment is expressed using the following notations:

$$\begin{aligned} T_3^{(j)} &= \text{tr} \left[(A^{(j)})^3 \right], \quad S_3^{(j)} = \sum_k (A_{kk}^{(j)})^3, \quad U^{(j)} = \sum_{kl} (A_{kl}^{(j)})^3, \\ R^{(j)} &= \sum_{kl} A_{kk}^{(j)} (A_{kl}^{(j)})^2 = \left[\text{diag}(A^{(j)}) \right]' \text{diag} \left[(A^{(j)})^2 \right], \\ B^{(j)} &= \sum_{kl} A_{kk}^{(j)} A_{kl}^{(j)} A_{ll}^{(j)} = \left[\text{diag}(A^{(j)}) \right]' A^{(j)} \text{diag}(A^{(j)}). \end{aligned}$$

Theorem 3.9. *The third moment of $T = \sum_{k,l=1}^n \prod_{j=1}^p A_{kl}^{(j)}$, where $A^{(j)}$ is doubly-centered, is given by*

$$\begin{aligned}
(n!)^{p-1} \mathbb{E}_P(T^3) &= [(n-1)!]^{p-1} \prod_{j=1}^p S_3^{(j)} + 4[(n-2)!]^{p-1} \prod_{j=1}^p [-S_3^{(j)} + U^{(j)}] \\
&+ 3[(n-2)!]^{p-1} \prod_{j=1}^p [T^{(j)} S_2^{(j)} - S_3^{(j)}] + 6[(n-2)!]^{p-1} \prod_{j=1}^p [-S_3^{(j)}] \\
&+ 12[(n-2)!]^{p-1} \prod_{j=1}^p [-S_3^{(j)} + R^{(j)}] + 6[(n-2)!]^{p-1} \prod_{j=1}^p [-S_3^{(j)} + B^{(j)}] \\
&+ 3[(n-3)!]^{p-1} \prod_{j=1}^p [-T^{(j)} S_2^{(j)} + 2S_3^{(j)}] + [(n-3)!]^{p-1} \prod_{j=1}^p [(T^{(j)})^3 - 3T^{(j)} S_2^{(j)} + 2S_3^{(j)}] \\
&+ 12[(n-3)!]^{p-1} \prod_{j=1}^p [-T^{(j)} S_2^{(j)} + 2S_3^{(j)} - B^{(j)}] + 12[(n-3)!]^{p-1} \prod_{j=1}^p [2S_3^{(j)} - R^{(j)}] \\
&+ 6[(n-3)!]^{p-1} \prod_{j=1}^p [T^{(j)}(T_2^{(j)} - S_2^{(j)}) + 2S_3^{(j)} - 2R^{(j)}] \\
&+ 24[(n-3)!]^{p-1} \prod_{j=1}^p [2S_3^{(j)} - R^{(j)} - B^{(j)}] + 24[(n-3)!]^{p-1} \prod_{j=1}^p [2S_3^{(j)} - U^{(j)} - R^{(j)}] \\
&+ 8[(n-3)!]^{p-1} \prod_{j=1}^p [T_3^{(j)} + 2S_3^{(j)} - 3R^{(j)}] \\
&+ 12[(n-4)!]^{p-1} \prod_{j=1}^p [T^{(j)} S_2^{(j)} - 6S_3^{(j)} + 2R^{(j)} + 2B^{(j)}] \\
&+ 6[(n-4)!]^{p-1} \prod_{j=1}^p [T^{(j)}(-T_2^{(j)} + S_2^{(j)}) - 6S_3^{(j)} + 2U^{(j)} + 4R^{(j)}] \\
&+ 3[(n-4)!]^{p-1} \prod_{j=1}^p [-(T^{(j)})^3 + 5T^{(j)} S_2^{(j)} - 6S_3^{(j)} + 2B^{(j)}] \\
&+ 12[(n-4)!]^{p-1} \prod_{j=1}^p [T^{(j)}(-T_2^{(j)} + 2S_2^{(j)}) - 6S_3^{(j)} + 3R^{(j)} + 2B^{(j)}] \\
&+ 8[(n-4)!]^{p-1} \prod_{j=1}^p [-6S_3^{(j)} + 2U^{(j)} + 3R^{(j)}] \\
&+ 24[(n-4)!]^{p-1} \prod_{j=1}^p [-T_3^{(j)} - 6S_3^{(j)} + U^{(j)} + 5R^{(j)} + B^{(j)}] \\
&+ 3[(n-5)!]^{p-1} \prod_{j=1}^p [(T^{(j)})^3 + 2T^{(j)}(T_2^{(j)} - 5S_2^{(j)}) + 24S_3^{(j)} - 8R^{(j)} - 8B^{(j)}] \\
&+ 12[(n-5)!]^{p-1} \prod_{j=1}^p [T^{(j)}(T_2^{(j)} - 2S_2^{(j)}) + 2T_3^{(j)} + 24S_3^{(j)} - 4U^{(j)} - 16R^{(j)} - 4B^{(j)}]
\end{aligned}$$

$$+[(n-6)!]^{p-1} \prod_{j=1}^p \left[-(T^{(j)})^3 - 6T^{(j)}(T_2^{(j)} - 3S_2^{(j)}) - 8T_3^{(j)} - 120S_3^{(j)} + 16U^{(j)} + 72R^{(j)} + 24B^{(j)} \right].$$

The principle of derivation remains the same for the terms involved in the formula of the third moment. However, they require more attention and are tedious. Then, with the same notation as in Theorem 3.9, four lemmas are needed for the formula of the third moment.

Lemma 3.3 (Sums for one or two distinct indices). *The following expressions hold*

$$\begin{aligned} \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^3 &= (n-1)! S_3^{(a)}, \\ \sum_{\sigma_a} [A_{ij}^{(a)}(\sigma_a)]^3 &= (n-2)! [-S_3^{(a)} + U^{(a)}], \\ \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{jj}^{(a)}(\sigma_a) &= (n-2)! [T^{(a)} S_2^{(a)} - S_3^{(a)}], \\ \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{ij}^{(a)}(\sigma_a) &= (n-2)! [-S_3^{(a)}], \\ \sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) [A_{ij}^{(a)}(\sigma_a)]^2 &= (n-2)! [-S_3^{(a)} + R^{(a)}], \\ \sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) &= (n-2)! [-S_3^{(a)} + B^{(a)}]. \end{aligned}$$

PROOF OF LEMMA 3.3. For two distinct indices i, j , the sums over all permutations are

$$\begin{aligned} \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^3 &= (n-1)! \sum_i (A_{ii}^{(a)})^3 \\ &= (n-1)! S_3^{(a)}, \\ \sum_{\sigma_a} [A_{ij}^{(a)}(\sigma_a)]^3 &= (n-2)! \sum'_{ij} (A_{ij}^{(a)})^3 \\ &= (n-2)! \left[\sum_{ij} (A_{ij}^{(a)})^3 - \sum_i (A_{ii}^{(a)})^3 \right] \\ &= (n-2)! [-S_3^{(a)} + U^{(a)}], \\ \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{jj}^{(a)}(\sigma_a) &= (n-2)! \sum'_{ij} (A_{ii}^{(a)})^2 A_{jj}^{(a)} \\ &= (n-2)! \left[\sum_i (A_{ii}^{(a)})^2 \sum_j A_{jj}^{(a)} - \sum_i (A_{ii}^{(a)})^3 \right] \\ &= (n-2)! [T^{(a)} S_2^{(a)} - S_3^{(a)}], \end{aligned}$$

$$\begin{aligned}
\sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{ij}^{(a)}(\sigma_a) &= (n-2)! \sum'_{ij} (A_{ii}^{(a)})^2 A_{ij}^{(a)} \\
&= (n-2)! \left[\sum_i (A_{ii}^{(a)})^2 \sum_j A_{ij}^{(a)} - \sum_i (A_{ii}^{(a)})^3 \right] \\
&= (n-2)! [-S_3^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) [A_{ij}^{(a)}(\sigma_a)]^2 &= (n-2)! \sum'_{ij} A_{ii}^{(a)} (A_{ij}^{(a)})^2 \\
&= (n-2)! \left[\sum_{ij} A_{ii}^{(a)} (A_{ij}^{(a)})^2 - \sum_i (A_{ii}^{(a)})^3 \right] \\
&= (n-2)! [-S_3^{(a)} + R^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) &= (n-2)! \sum'_{ij} A_{ii}^{(a)} A_{jj}^{(a)} A_{ij}^{(a)} \\
&= (n-2)! \left[\sum_{ij} A_{ii}^{(a)} A_{ij}^{(a)} A_{ii}^{(a)} - \sum_i (A_{ii}^{(a)})^3 \right] \\
&= (n-2)! [-S_3^{(a)} + B^{(a)}].
\end{aligned}$$

□

Lemma 3.4 (Sums for three distinct indices). *The following expressions hold*

$$\begin{aligned}
\sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{jk}^{(a)}(\sigma_a) &= (n-3)! [-T^{(a)} S_2^{(a)} + 2S_3^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) A_{kk}^{(a)}(\sigma_a) &= (n-3)! [(T^{(a)})^3 - 3T^{(a)} S_2^{(a)} + 2S_3^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) &= (n-3)! [-T^{(a)} S_2^{(a)} + 2S_3^{(a)} - B^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) &= (n-3)! [2S_3^{(a)} - R^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) [A_{jk}^{(a)}(\sigma_a)]^2 &= (n-3)! [T^{(a)} (T_2^{(a)} - S_2^{(a)}) + 2S_3^{(a)} - 2R^{(a)}], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) A_{jk}^{(a)}(\sigma_a) &= (n-3)! [2S_3^{(a)} - R^{(a)} - B^{(a)}], \\
\sum_{\sigma_a} [A_{ij}^{(a)}(\sigma_a)]^2 A_{ik}^{(a)}(\sigma_a) &= (n-3)! [2S_3^{(a)} - U^{(a)} - R^{(a)}], \\
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) A_{jk}^{(a)}(\sigma_a) &= (n-3)! [T_3^{(a)} + 2S_3^{(a)} - 3R^{(a)}].
\end{aligned}$$

Lemma 3.5 (Sums for four distinct indices). *The following expressions hold*

$$\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) = (n-4)! [T^{(a)} S_2^{(a)} - 6S_3^{(a)} + 2R^{(a)} + 2B^{(a)}],$$

$$\begin{aligned}
\sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^2 A_{kl}^{(a)}(\sigma_a) &= (n-4)! \left[T^{(a)}(-T_2^{(a)} + S_2^{(a)}) - 6S_3^{(a)} + 2U^{(a)} + 4R^{(a)} \right], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jj}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) &= (n-4)! \left[-(T^{(a)})^3 + 5T^{(a)}S_2^{(a)} - 6S_3^{(a)} + 2B^{(a)} \right], \\
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jk}^{(a)}(\sigma_a) A_{jl}^{(a)}(\sigma_a) &= (n-4)! \left[T^{(a)}(-T_2^{(a)} + 2S_2^{(a)}) - 6S_3^{(a)} + 3R^{(a)} + 2B^{(a)} \right], \\
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) A_{il}^{(a)}(\sigma) &= (n-4)! \left[-6S_3^{(a)} + 2U^{(a)} + 3R^{(a)} \right], \\
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) A_{jl}^{(a)}(\sigma_a) &= (n-4)! \left[-T_3^{(a)} - 6S_3^{(a)} + U^{(a)} + 5R^{(a)} + B^{(a)} \right].
\end{aligned}$$

Lemma 3.6 (Sums for five and six distinct indices). *The following expressions hold*

$$\begin{aligned}
\sum_{\sigma_a} A_{ii}^{(a)}(\sigma_a) A_{jk}^{(a)}(\sigma_a) A_{lm}^{(a)}(\sigma_a) &= (n-5)! \left[(T^{(a)})^3 + 2T^{(a)}(T_2^{(a)} - 5S_2^{(a)}) \right. \\
&\quad \left. + 24S_3^{(a)} - 8R^{(a)} - 8B^{(a)} \right], \\
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{ik}^{(a)}(\sigma_a) A_{lm}^{(a)}(\sigma_a) &= (n-5)! \left[T^{(a)}(T_2^{(a)} - 2S_2^{(a)}) + 2T_3^{(a)} \right. \\
&\quad \left. + 24S_3^{(a)} - 4U^{(a)} - 16R^{(a)} - 4B^{(a)} \right], \\
\sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) A_{mh}^{(a)}(\sigma) &= (n-6)! \left[-(T^{(a)})^3 - 6T^{(a)}(T_2^{(a)} - 3S_2^{(a)}) - 8T_3^{(a)} \right. \\
&\quad \left. - 120S_3^{(a)} + 16U^{(a)} + 72R^{(a)} + 24B^{(a)} \right].
\end{aligned}$$

PROOF OF LEMMAS 3.4, 3.5 AND 3.6. The proof is similar to the proof of the Lemma 3.3. \square

PROOF OF THEOREM 3.9.

$$\begin{aligned}
&n! \mathbb{E}_P T^3 \\
&= \sum_{ijklmh} \prod_{a=1}^p \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a) A_{kl}^{(a)}(\sigma_a) A_{mh}^{(a)}(\sigma_a) \\
&= \sum_i \prod_{a=1}^p \sum_{\sigma_a} [A_{ii}^{(a)}(\sigma_a)]^3 \\
&\quad + \sum'_{ij} \prod_{a=1}^p \sum_{\sigma_a} \left[4[A_{ij}^{(a)}(\sigma_a)]^3 + 3[A_{ii}^{(a)}(\sigma_a)]^2 A_{jj}^{(a)}(\sigma_a) + 6[A_{ii}^{(a)}(\sigma_a)]^2 A_{ij}^{(a)}(\sigma_a) \right. \\
&\quad \left. + 12A_{ii}^{(a)}(\sigma_a)[A_{ij}^{(a)}(\sigma_a)]^2 + 6A_{ii}^{(a)}(\sigma_a)A_{jj}^{(a)}(\sigma_a)A_{ij}^{(a)}(\sigma_a) \right] \\
&\quad + \sum'_{ijk} \prod_{a=1}^p \sum_{\sigma_a} \left[3[A_{ii}^{(a)}(\sigma_a)]^2 A_{jk}^{(a)}(\sigma_a) + A_{ii}^{(a)}(\sigma_a)A_{jj}^{(a)}(\sigma_a)A_{kk}^{(a)}(\sigma_a) \right. \\
&\quad \left. + 12A_{ii}^{(a)}(\sigma_a)A_{jj}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a) + 12A_{ii}^{(a)}(\sigma_a)A_{ij}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a) + 6A_{ii}^{(a)}(\sigma_a)[A_{jk}^{(a)}(\sigma_a)]^2 \right]
\end{aligned}$$

$$\begin{aligned}
& + 24A_{ii}^{(a)}(\sigma_a)A_{ij}^{(a)}(\sigma_a)A_{jk}^{(a)}(\sigma_a) + 24[A_{ij}^{(a)}(\sigma_a)]^2A_{ik}^{(a)}(\sigma_a) + 8A_{ij}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a)A_{jk}^{(a)}(\sigma_a) \\
& + \sum'_{ijkl} \prod_{a=1}^p \sum_{\sigma_a} \left[12A_{ii}^{(a)}(\sigma_a)A_{ij}^{(a)}(\sigma_a)A_{kl}^{(a)}(\sigma_a) + 6[A_{ij}^{(a)}(\sigma_a)]^2A_{kl}^{(a)}(\sigma_a) \right. \\
& + 3A_{ii}^{(a)}(\sigma_a)A_{jj}^{(a)}(\sigma_a)A_{kl}^{(a)}(\sigma_a) + 12A_{ii}^{(a)}(\sigma_a)A_{jk}^{(a)}(\sigma_a)A_{jl}^{(a)}(\sigma_a) \\
& + 8A_{ij}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a)A_{il}^{(a)}(\sigma_a) + 24A_{ij}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a)A_{jl}^{(a)}(\sigma_a) \left. \right] \\
& + \sum'_{ijklm} \prod_{a=1}^p \sum_{\sigma_a} \left[3A_{ii}^{(a)}(\sigma_a)A_{jk}^{(a)}(\sigma_a)A_{lm}^{(a)}(\sigma_a) + 12A_{ij}^{(a)}(\sigma_a)A_{ik}^{(a)}(\sigma_a)A_{lm}^{(a)}(\sigma_a) \right] \\
& + \sum'_{ijklmh} \prod_{a=1}^p \sum_{\sigma_a} A_{ij}^{(a)}(\sigma_a)A_{kl}^{(a)}(\sigma_a)A_{mh}^{(a)}(\sigma_a). \tag{3.20}
\end{aligned}$$

Sums over permutations in Lemmas 3.3, 3.4, 3.5 and 3.6 do not depend on the values of indices, but only on the number of different indices selected. From Equation (3.20) and Lemmas 3.3, 3.4, 3.5 and 3.6, the third moment in Theorem 3.9 follows. \square

BIBLIOGRAPHY

- R. Beran, M. Bilodeau, and P. Lafaye de Micheaux. Nonparametric tests of independence between random vectors. *Journal of Multivariate Analysis*, 98 (9):1805–1824, 2007.
- M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. Springer Texts in Statistics. Springer-Verlag, New York, 1999.
- M. Bilodeau and P. Lafaye de Micheaux. A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, 95(2):345–369, 2005.
- M. Bilodeau and P. Lafaye de Micheaux. A-dependence statistics for mutual and serial independence of categorical variables. *Journal of Statistical Planning and Inference*, 139(7):2407–2419, 2009.
- J.R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32(2):485–498, 1961.
- D.S. Cotterill and M. Csörgő. On the limiting distribution of and critical values for the multivariate Cramér-von Mises statistic. *The Annals of Statistics*, 10 (1):233–244, 1982.
- D.S. Cotterill and M. Csörgő. On the limiting distribution of and critical values for the Hoeffding, Blum, Kiefer, Rosenblatt independence criterion. *Statistics & Decisions. International Journal for Statistical Theory and Related Fields*, 3 (1-2):1–48, 1985.

- S. Csörgő. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(2):203–229, 1981.
- S. Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16(3):290–299, 1985.
- P. Deheuvels. An asymptotic decomposition for multivariate distribution-free tests of independence. *Journal of Multivariate Analysis*, 11(1):102–113, 1981.
- P. Duchesne, K. Ghoudi, and B. Rémillard. On testing for independence between the innovations of several time series. *The Canadian Journal of Statistics*, 40(3):447–479, 2012.
- Y. Escoufier. Le traitement des variables vectorielles. *Biometrics*, 29(4):751–760, 1973.
- Y. Fan, P. Lafaye de Micheaux, S. Penev, and Salopek D. Multivariate nonparametric test of independence. *Bernoulli (In revision)*, 2015.
- T.S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Taylor & Francis, 1996.
- C. Genest and B. Rémillard. Tests of independence or randomness based on the empirical copula process. *Test*, 13(2):335–370, 2005.
- K. Ghoudi, R.J. Kulperger, and B. Rémillard. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79(2):191–218, 2001.
- A. Gretton, O. Bousquet, A. Smola, and B. Scholköpfung. Measuring statistical dependence with Hilbert-Schmidt norms. In S. Jain, H. Simon, and E. Tomita, editors, *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2005.
- A. Gretton, K. Fukumizu, and B.K. Sriperumbudur. Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1285–1294, 12 2009.
- N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617, 1990.
- S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics. Theory and Applications*, 6(2):65–70, 1979.
- H. Joe. Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10), 2006.
- J. Josse and S. Holmes. Tests of independence and beyond. *arXiv:1307.7383v3*, 2014.
- J. Josse, J. Pagès, and F. Husson. Testing the significance of the rv coefficient. *Computational Statistics & Data Analysis*, 53(1):82–91, 2008.

- O. Kallenberg. *Foundations of Modern Probability*. Probability and its applications. Springer, New York, 2002.
- A. Kankainen. *Consistent Testing of Total Independence Based on the Empirical Characteristic Function*. PhD thesis, University of Jyväskylä, 1995.
- F. Kazi-Aoual, S. Hitier, R. Sabatier, and J.-D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics & Data Analysis*, 20(6):643–656, 1995.
- J. Kellermeier. The empirical characteristic function and large sample hypothesis testing. *Journal of Multivariate Analysis*, 10(1):78–87, 1980.
- I. Kojadinovic and M. Holmes. Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *Journal of Multivariate Analysis*, 100(6):1137–1154, 2009.
- I. Kojadinovic and J. Yan. Tests of serial independence for continuous multivariate time series based on a Möbius decomposition of the independence empirical copula process. *Annals of the Institute of Statistical Mathematics*, 63(2):347–373, 2011.
- S. Korkmaz, D. Goksuluk, and G. Zararsiz. MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.
- H.O. Lancaster. Complex contingency tables treated by the partition of χ^2 . *Journal of the Royal Statistical Society. Series B. Methodological*, 13(2):242–249, 1951.
- R.C. Littell and J.L. Folks. Asymptotic optimality of Fisher’s method of combining independent tests. *Journal of the American Statistical Association*, 66(336):802–806, 1971.
- R.C. Littell and J.L. Folks. Asymptotic optimality of Fisher’s method of combining independent tests II. *Journal of the American Statistical Association*, 68(341):193–194, 1973.
- R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284–3305, 2013.
- C. Minas, E. Curry, and G. Montana. A distance-based test of association between paired heterogeneous genomic data. *Bioinformatics*, 29(20):2555–2563, 2013.
- A.C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, New York, 1995.
- J. Romano and A. Siegel. *Counterexamples in Probability and Statistics*. Wadsworth, London, 1986.
- D. Sejdinovic, A. Gretton, and W. Bergsma. A kernel test for three-variable interactions. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*

- 26, pages 1124–1132. Curran Associates, Inc., 2013a.
- D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291, 2013b.
- G.J. Székely and M.L. Rizzo. Rejoinder: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1303–1308, 2009.
- G.J. Székely, M.L. Rizzo, and N.K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.
- D.E. Tyler. Robustness and efficiency properties of scatter matrices. *Biometrika*, 70(2):411–420, 1983.
- W. Whitt. Weak convergence of probability measures on the function space $C[0, \infty)$. *The Annals of Mathematical Statistics*, 41(3):939–944, 1970.
- K. Zhang, J. Peters, D. Janzing, and B. Schölkopf. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 804–813. AUAI Press, Corvallis, Oregon, 2011.

Chapitre 4

CONCLUSION

Cette thèse a traité des tests d'indépendance non paramétriques entre deux ou plusieurs vecteurs aléatoires. Elle s'est aussi intéressée à l'indépendance sérielle d'une suite multidimensionnelle stationnaire. L'originalité de cette thèse est une estimation très rapide et précise de la valeur- p des tests considérés. Cette estimation des valeurs- p donne de meilleurs résultats que d'autres méthodes proposées auparavant dans la littérature, particulièrement en grandes dimensions ou en petites tailles échantillonnelles. Nous avons aussi généralisé les notions de distance de covariance et de critère d'indépendance de Hilbert-Schmidt (*HSIC*) à l'indépendance mutuelle entre plusieurs vecteurs de même qu'à l'indépendance sérielle d'une suite multidimensionnelle stationnaire. Une méthode prometteuse de sélection adaptative des paramètres d'échelle du noyau caractéristique du test *HSIC* a aussi été proposée.

Dans le chapitre 2, les tests d'indépendance entre deux vecteurs aléatoires ont été traités. Un résultat simple mais important établit l'équivalence, en terme de puissance, entre la distance de covariance et le test *HSIC* avec pour noyau la fonction caractéristique d'une distribution stable dans laquelle les paramètres d'échelle sont suffisamment petits. Cette équivalence stipule qu'il est toujours possible de trouver un test *HSIC* dont la puissance est, à toute fin pratique, la même que celle d'un test de distance de covariance. Un test de distance de covariance nécessite seulement la spécification d'un ordre α par rapport au test *HSIC* qui requiert aussi la spécification des paramètres d'échelle. Toutefois, une méthode adaptative de sélection des paramètres d'échelle a été proposée. Dans trois simulations, dont deux empruntées de la littérature en apprentissage automatique, le test *HSIC* avec cette méthode adaptative de sélection est plus puissant qu'un test de distance de covariance.

Tous les tests obtenus prennent la forme d'un produit scalaire entre deux matrices doublement centrées. La détermination rapide et précise de la valeur- p des

tests a reposé sur une approximation de la distribution exacte de permutation par une distribution continue de Pearson de type III. Cette approximation nécessite les trois premiers moments de la distribution exacte de permutation. En revanche, et c'est là où réside son grand avantage, elle ne requiert aucun rééchantillonnage. Gretton et al. (2008) ont estimé les deux premiers moments de la distribution asymptotique du test *HSIC* et ont utilisé une distribution gamma comme approximation. Des simulations ont montré que leur approche est souvent imprécise, même lorsque la taille de l'échantillon est égale à 100. La distribution de Pearson de type III présente un grand avantage d'inclure une correction par le coefficient d'asymétrie.

Le chapitre 3 a généralisé les tests d'indépendance entre deux vecteurs aux tests d'indépendance mutuelle entre plusieurs vecteurs. Il a aussi traité des tests d'indépendance sérielle d'une suite multidimensionnelle stationnaire. La décomposition de Möbius du processus d'indépendance par les fonctions caractéristiques a permis de définir des statistiques de dépendance de type Cramér-von Mises pour chaque sous-ensemble de deux ou plusieurs vecteurs. Il y est démontré que ces statistiques s'écrivent comme une somme sur tous les éléments d'une matrice exprimée comme un produit, composante par composante, de matrices doublement centrées. Des tests généralisés basés sur le critère d'indépendance de Hilbert-Schmidt et sur la distance de covariance en sont déduits. Tout comme dans le chapitre 2, une équivalence est établie entre le test basé sur la distance de covariance et le test *HSIC* de noyau caractéristique d'une distribution stable avec des paramètres d'échelle suffisamment petits. La méthode de sélection adaptative des paramètres d'échelle du noyau caractéristique a également été modifiée pour les tests d'indépendance mutuelle et sérielle. Une simulation a permis de montrer qu'elle améliore la puissance lorsque la distribution des observations est une copule de Student.

Un résultat fascinant est l'obtention des trois premiers moments exacts de la distribution de permutation des statistiques de dépendance. Ce résultat a nécessité un travail rigoureux et minutieux. Le succès qu'a connu la distribution de Pearson de type III comme estimation de la distribution exacte de permutation des statistiques a été mis en valeur. Ainsi, toutes les valeurs- p des tests développés ont été estimées par une distribution de Pearson de type III. Les tests présentés dans cette thèse sont applicables en petites comme en grandes dimensions. Une combinaison des valeurs- p à la Fisher a aussi été mise de l'avant dans un test global d'indépendance mutuelle avec un bon contrôle du taux d'erreur de première espèce. Une méthodologie similaire a été développée pour le test d'indépendance sérielle d'une suite. Des applications à des données réelles environnementales et

financières ont mis en exergue l'utilité des tests développés. Spécifiquement, pour les taux de croissance des séries financières journalières S&P/TSX composite, S&P500 et DOW JONES, prises à la clôture et observées sur la période allant du 2 janvier 2014 au 2 mars 2016, les tests ont montré qu'il y a une dépendance entre la croissance à une journée donnée et celles des quatre journées précédentes.

Dans les tests d'indépendance mutuelle et sérielle, les propriétés asymptotiques du processus expérimental d'indépendance par les fonctions caractéristiques à la base du test *HSIC* et de la distance de covariance ont été démontrées. En revanche, pour les fonctionnelles définissant les statistiques de dépendance, la distribution asymptotique a été démontrée uniquement pour les tests *HSIC*. Pour la distance de covariance, la difficulté engendrée par l'utilisation d'une mesure non intégrable a simplement été expliquée.

Tous les résultats de cette thèse ont porté sur des vecteurs aléatoires, c'est-à-dire des éléments aléatoires dans des espaces euclidiens. Une avenue de recherche que j'aimerais explorer par la suite est l'étude de tests d'indépendance pour des éléments aléatoires dans des espaces métriques. Par exemple, les données en neurosciences sont souvent sous la forme d'images ou de fonctions. Les résultats de Lyons (2013) permettent la généralisation aux espaces métriques en présence de seulement deux éléments aléatoires. La généralisation à l'indépendance mutuelle entre plusieurs éléments aléatoires ou à l'indépendance sérielle d'une suite stationnaire d'éléments aléatoires est un défi de taille. Une avenue de recherche que j'aimerais également explorer est le choix adaptatif de l'indice α .

BIBLIOGRAPHIE

- A. Gretton, F. Kenji, C.-H. Teo, L. Song, B. Schölkopf, et A. Smola. A kernel statistical test of independence. In J.C. Platt, D. Koller, Y. Singer, et S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 585–592. Curran Associates, Inc., 2008.
- R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5) :3284–3305, 2013.

Annexe A

PROGRAMME R ET C++ DU PREMIER ARTICLE

```
setwd("C:/Users/AURELBRELL/Desktop/rcpp")
system("R CMD SHLIB dcovp.cpp -o dcovp.dll") # pour compiler
# le programme dcovp.cpp

## Fonction de calcul d'une matrice doublement centrée
fctA <- function(a){
  mean.ak. <- rowMeans(a)
  mean.a.l <- colMeans(a)
  mean.a <- mean(a)
  A <- sweep(a, MARGIN = 1, STATS = mean.ak., FUN = "-")
  A <- sweep(A, MARGIN = 2, STATS = mean.a.l, FUN = "-")
  A <- A + mean.a
  return(A)
}

## Fonction de calcul des deux premiers moments de tr(A1A2)
perm.var.tr.A1A2 <- function(A1,A2){
  if(nrow(A1)!=ncol(A1) | nrow(A1)!=nrow(A2) | nrow(A1)!=ncol(A2) ){
    stop("The two matrix must be square and have the same dimension")
  }
  n <- nrow(A1)
  # Notations et fonctions de base
  G <- function(A){
    dA <- diag(A)
    T <- sum(dA)
    A2c <- A^2
```



```

    T2 <- sum(A2c)
    dA2 <- dA^2
    S2 <- sum(dA2)
    A2m <- A%*%A
    list(T=T,T2=T2,S2=S2)
  }
  T <- c()
  T2 <- c() ; S2 <- c()
  out1 <- G(A1) ; out2 <- G(A2)
  T[1] <- out1$T ; T[2] <- out2$T
  T2[1] <- out1$T2 ; T2[2] <- out2$T2
  S2[1] <- out1$S2 ; S2[2] <- out2$S2
  # Formule du 1er moment
  mom1 <- prod(T)/n + prod(-T)/(n*(n-1))
  # Formule du 2e moment
  mom2 <- prod(S2)/n+( prod(T^2-S2)+2*prod(T2-S2)+4*prod(-S2) )/(n*(n-1));
          + ( 4*prod(2*S2-T2)+2*prod(2*S2-T^2) )/(n*(n-1)*(n-2));
          + prod(2*T2-6*S2+T^2)/(n*(n-1)*(n-2)*(n-3))
  mean.T <- mom1
  variance.T <- (mom2 - mom1^2)
  return(list(mean.T=mean.T,variance.T=variance.T))
}

## Fonction de calcul des trois premiers moments de tr(A1A2)
perm.mom.tr.A1A2 <- function(A1,A2){
  if(nrow(A1)!=ncol(A1) | nrow(A1)!=nrow(A2) | nrow(A1)!=ncol(A2) ){
    stop("The two matrix must be square and have the same dimension")
  }
  n <- nrow(A1)
  # Notations et fonctions de base
  G <- function(A){
    dA <- diag(A)
    T <- sum(dA)
    A2c <- A^2
    T2 <- sum(A2c)
    dA2 <- dA^2
    S2 <- sum(dA2)
    A2m <- A%*%A

```

```

T3 <- sum(A2m*A)
S3 <- sum(dA2*dA)
U <- sum(A2c*A)
R <- as.vector(t(dA)%*%diag(A2m))
B <- as.vector(t(dA)%*%A%*%dA)
list(T=T,T2=T2,S2=S2,T3=T3,S3=S3,U=U,R=R,B=B)
}

T <- c()
T2 <- c() ; S2 <- c()
T3 <- c() ; S3 <- c() ; U <- c() ; R <- c() ; B <- c()
out1 <- G(A1) ; out2 <- G(A2)
T[1] <- out1$T ; T[2] <- out2$T
T2[1] <- out1$T2 ; T2[2] <- out2$T2
S2[1] <- out1$S2 ; S2[2] <- out2$S2
T3[1] <- out1$T3 ; T3[2] <- out2$T3
S3[1] <- out1$S3 ; S3[2] <- out2$S3
U[1] <- out1$U ; U[2] <- out2$U
R[1] <- out1$R ; R[2] <- out2$R
B[1] <- out1$B ; B[2] <- out2$B
# Formule du 1er moment
mom1 <- prod(T)/n + prod(-T)/(n*(n-1))
# Formule du 2e moment
mom2 <- prod(S2)/n+(prod(T^2-S2)+2*prod(T2-S2)+4*prod(-S2))/(n*(n-1));
+ ( 4*prod(2*S2-T2)+2*prod(2*S2-T^2) )/(n*(n-1)*(n-2));
+ prod(2*T2-6*S2+T^2)/(n*(n-1)*(n-2)*(n-3))
# Formule du 3e moment
SP1 <- prod(S3)/n
SP2 <- ( 4*prod(-S3+U)+3*prod(T*S2-S3) + 6*prod(-S3);
+ 12*prod(-S3+R) + 6*prod(-S3+B) )/(n*(n-1))
SP3 <- ( 3*prod(-T*S2+2*S3) + prod(T^3-3*T*S2+2*S3);
+ 12*prod(-T*S2+2*S3-B) + 12*prod(2*S3-R) + 24*prod(2*S3-R-B);
+ 6*prod(T*(T2-S2)+2*S3-2*R) + 24*prod(2*S3-U-R);
+ 8*prod(T3+2*S3-3*R) )/(n*(n-1)*(n-2))
SP4 <- ( 12*prod(T*S2-6*S3+2*R+2*B) + 6*prod(T*(-T2+S2)-6*S3+2*U+4*R);
+ 3*prod(-T^3+5*T*S2-6*S3+2*B) + 12*prod(T*(-T2+2*S2)-6*S3+3*R+2*B);
+ 8*prod(-6*S3+2*U+3*R) + 24*prod(-T3-6*S3+U+5*R+B) );
/(n*(n-1)*(n-2)*(n-3))
SP5 <- ( 3*prod(T^3+2*T*(T2-5*S2)+24*S3-8*R-8*B);

```

```

      + 12*prod(T*(T2-2*S2)+2*T3+24*S3-4*U-16*R-4*B) );
      /(n*(n-1)*(n-2)*(n-3)*(n-4))
SP6 <- prod(-T^3-6*T*(T2-3*S2)-8*T3-120*S3+16*U+72*R+24*B);
      /(n*(n-1)*(n-2)*(n-3)*(n-4)*(n-5))
mom3 <- SP1+SP2+SP3+SP4+SP5+SP6
mean.T <- mom1
variance.T <- (mom2 - mom1^2)
skewness.T <- (mom3 - 3 * variance.T * mom1 - mom1^3)/(variance.T^(3/2))
return(list(mean.T=mean.T,variance.T=variance.T,skewness.T=skewness.T))
}

## Fonction de la statistique et du test dans R
indep.pearson <- function (X1, X2, index=1, cte=NULL, method="dcov ou hsic")
{
  if(is.vector(X1)){
    n=length(X1)
  }
  if(is.matrix(X1)){
    n=nrow(X1)
  }
  a1 <- as.matrix(dist(X1)^index)
  a2 <- as.matrix(dist(X2)^index)
  if(method=="hsic"){
    fcta.b <- function(a,b){
      beta <- b/median(a[a>0])
      a <- exp( -a*(beta^index) )
      return(a)
    }
    if (is.null(cte)){
      var.perm <- function(b){
        a1.b1 <- fcta.b(a1,b[1])
        a2.b2 <- fcta.b(a2,b[2])
        A1 <- fctA(a1.b1)
        A2 <- fctA(a2.b2)
        -perm.var.tr.A1A2(A1,A2)$variance.T
      }
      out.opt <- nlm(var.perm, p=c(.01/median((a1[a1>0])^(1/index)),;
        .01/median((a2[a2>0])^(1/index))))
    }
  }
}

```

```

b.opt <- out.opt$estimate
a1 <- fcta.b(a1,b.opt[1])
a2 <- fcta.b(a2,b.opt[2])
  }else{
a1 <- fcta.b(a1,cte[1])
a2 <- fcta.b(a2,cte[2])
  }
}
A1 <- fctA(a1)
A2 <- fctA(a2)
TrA1A2 <- sum(A1*A2)
out.perm <- perm.mom.tr.A1A2(A1,A2)
mean.T <- out.perm$mean.T
variance.T <- out.perm$variance.T
as <- out.perm$skewness.T
TrA1A2.std <- (TrA1A2 - mean.T)/(variance.T^0.5)
if (as >= 0)
  prob <- pgamma(TrA1A2.std - (-2/as), shape = (4/as^2),,
    scale = (as/2), lower.tail = FALSE)
if (as < 0)
  prob = pgamma(as/abs(as) * TrA1A2.std + 2/abs(as),,
    shape = (4/as^2), scale = (abs(as)/2))
return(list(stat = TrA1A2/(n^2), mean.T = mean.T,,
  variance.T = variance.T, skewness.T = as, p.value = prob))
}

```

```

## Fonctions écrites en C++ et contenues dans le fichier dcovp.cpp

```

```

#include <iostream>
using namespace std;
#include <R.h>
#include <Rmath.h>
#include <R_ext/Rdynload.h>

```

```

/* ## compute a zero matrix with r rows and c columns */

```

```

double *alloc_matrix(int r, int c){
  int i,j;
  double *a;

```

```

    a = new double[r*c];
    for (i=0; i<r; i++) {
        for (j=0; j<i; j++) {
            a[i*c+i] = 0.0;
            a[i*c+j] = 0.0;
            a[j*c+i] = 0.0;
        }
    }
    return a;
}

/* ## Fonction for euclidian matrix */

double *Eucl_distance(double *x, int n, int d, double index){
    /*
    interpret x as an d by n matrix, so transpose the n by d matrix in R
    compute the index-Euclidean distance matrix
    */
    int i, j, k, p, q;
    double dsum, dif, *a;
    a = alloc_matrix(n,n);
    for (i=0; i<n; i++) {
        p = i*d;
        for (j=0; j<i; j++) {
            dsum = 0.0;
            q = j*d;
            for (k=0; k<d; k++) {
                dif = *(x+p+k) - *(x+q+k);
                dsum = dsum + dif*dif;
            }
            a[i*n+j] = a[j*n+i] = pow(sqrt(dsum),index);
        }
    }
    return a;
}

/* ##### Function for double centered matrix */

```

```

double *fctA(double *a, int n){
    /* -computes the A_{kl} distance from the distance matrix (a_{kl}) */
    int i, j;
    double meana=0.0;
    double *meanak, *meanal;
    meanak = new double[n];
    meanal = new double[n];
    for (i=1 ; i<=n ; i++) {
        meanak[i-1] = 0.0;
        meanal[i-1] = 0.0;
    }
    for (i=1 ; i<=n ; i++) {
        for (j=1 ; j<=n ; j++) {
            meanak[i-1] = meanak[i-1] + a[(j-1)*n+(i-1)];
            meanal[i-1] = meanal[i-1] + a[(i-1)*n+(j-1)];
            meana = meana + a[(j-1)*n+(i-1)];
        }
        meanak[i-1] = meanak[i-1]/n;
        meanal[i-1] = meanal[i-1]/n;
    }
    meana = meana/n/n;
    for (i=1 ; i<=n ; i++) {
        for (j=1 ; j<=n ; j++) {
            a[(j-1)*n+(i-1)] = a[(j-1)*n+(i-1)]-meanak[i-1]-meanal[j-1]+meana;
        }
    }
    return a;
}

/* ##### Function of statistic (main function) */
extern "C" {
void dcovp(double *x1, double *x2, double *stat, double *mom,
    double *variance, double *skewness, int *nrowx, int *ncolx1,
    int *ncolx2, double *index, char method){
/* interpret A1 and A2 as an n by n matrix, in row order (n vectors in R^n)
    compute the 1st three permutation moments of tr(A1A2) */
// initialisation
    int n=nrowx[0], d1=ncolx1[0], d2=ncolx2[0], alpha=index[0], i, j, k;

```

```

double A1ii, A2ii, A1ij, A2ij, A1ik, A2ik, A1jj, A2jj, A1jk, A2jk;
double A1ii2, A2ii2, A1ij2, A2ij2;
double T11=0.0, T12=0.0;
double T21=0.0, T22=0.0, S21=0.0, S22=0.0, TrA1A2=0.0;
double T31=0.0, T32=0.0, S31=0.0, S32=0.0, U1=0.0, U2=0.0;
double R1=0.0, R2=0.0, B1=0.0, B2=0.0;
double m1, m2, m3, var, et, skew;
// distance matrix
double *a1, *a2;
a1 = Eucl_distance(x1, n, d1, alpha);
a2 = Eucl_distance(x2, n, d2, alpha);
// double-centering of a matrix (A_kl)
double *A1, *A2;
A1 = fctA(a1,n);
A2 = fctA(a2,n);
// compute the first three moments of tr(A1A2)
for(i=0; i<n; i++){
    A1ii = A1[i*n+i];
    A2ii = A2[i*n+i];
    T11 = T11 + A1ii;
    T12 = T12 + A2ii;
    A1ii2 = A1ii*A1ii;
    A2ii2 = A2ii*A2ii;
    S21 = S21 + A1ii2;
    S22 = S22 + A2ii2;
    S31 = S31 + A1ii2*A1ii;
    S32 = S32 + A2ii2*A2ii;
    for(j=0; j<n; j++){
        A1ij = A1[i*n+j];
        A2ij = A2[i*n+j];
        A1jj = A1[j*n+j];
        A2jj = A2[j*n+j];
        A1ij2 = A1ij*A1ij;
        A2ij2 = A2ij*A2ij;
        T21 = T21 + A1ij2;
        T22 = T22 + A2ij2;
        TrA1A2 = TrA1A2 + A1ij*A2ij;
        U1 = U1 + A1ij*A1ij2;
    }
}

```

```

U2 = U2 + A2ij*A2ij2;
R1 = R1 + A1ii*A1ij2;
R2 = R2 + A2ii*A2ij2;
B1 = B1 + A1ii*A1ij*A1jj;
B2 = B2 + A2ii*A2ij*A2jj;
for(k=0; k<n; k++){
    A1ik = A1[i*n+k];
    A2ik = A2[i*n+k];
    A1jk = A1[j*n+k];
    A2jk = A2[j*n+k];
    T31 = T31 + A1ij*A1jk*A1ik;
    T32 = T32 + A2ij*A2jk*A2ik;
}
}
}
stat[0] = TrA1A2/n/n;
m1 = T11*T12/(n-1);
mom[0] = m1;
m2 = S21*S22/n + (T11*T11-S21)*(T12*T12-S22)/n/(n-1);
m2 = m2 + 2*(T21-S21)*(T22-S22)/n/(n-1) + 4*S21*S22/n/(n-1);
m2 = m2 + 4*(-T21+2*S21)*(-T22+2*S22)/n/(n-1)/(n-2);
m2 = m2 + 2*(-T11*T11+2*S21)*(-T12*T12+2*S22)/n/(n-1)/(n-2);
m2 = m2 + (T11*T11+2*T21-6*S21)*(T12*T12+2*T22-6*S22);
/n/(n-1)/(n-2)/(n-3);
mom[1] = m2;
m3 = S31*S32/n + 4*(-S31+U1)*(-S32+U2)/n/(n-1);
m3 = m3 + 3*(T11*S21-S31)*(T12*S22-S32)/n/(n-1);
m3 = m3 + 6*(-S31)*(-S32)/n/(n-1);
m3 = m3 + 12*(-S31+R1)*(-S32+R2)/n/(n-1);
m3 = m3 + 6*(-S31+B1)*(-S32+B2)/n/(n-1);
m3 = m3 + 3*(-T11*S21+2*S31)*(-T12*S22+2*S32)/n/(n-1)/(n-2);
m3 = m3 + (T11*T11*T11-3*T11*S21+2*S31);
*(T12*T12*T12-3*T12*S22+2*S32)/n/(n-1)/(n-2);
m3 = m3 + 12*(-T11*S21+2*S31-B1);
*(-T12*S22+2*S32-B2)/n/(n-1)/(n-2);
m3 = m3 + 12*(2*S31-R1)*(2*S32-R2)/n/(n-1)/(n-2);
m3 = m3 + 6*(T11*(T21-S21)+2*S31-2*R1);
*(T12*(T22-S22)+2*S32-2*R2)/n/(n-1)/(n-2);

```



```

m3 = m3 + 24*(2*S31-R1-B1)*(2*S32-R2-B2)/n/(n-1)/(n-2);
m3 = m3 + 24*(2*S31-U1-R1)*(2*S32-U2-R2)/n/(n-1)/(n-2);
m3 = m3 + 8*(T31+2*S31-3*R1)*(T32+2*S32-3*R2)/n/(n-1)/(n-2);
m3 = m3 + 12*(T11*S21-6*S31+2*R1+2*B1);
*(T12*S22-6*S32+2*R2+2*B2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 6*(T11*(-T21+S21)-6*S31+2*U1+4*R1);
*(T12*(-T22+S22)-6*S32+2*U2+4*R2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 3*(-T11*T11*T11+5*T11*S21-6*S31+2*B1);
*(-T12*T12*T12+5*T12*S22-6*S32+2*B2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 12*(T11*(-T21+2*S21)-6*S31+3*R1+2*B1);
*(T12*(-T22+2*S22)-6*S32+3*R2+2*B2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 8*(-6*S31+2*U1+3*R1);
*(-6*S32+2*U2+3*R2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 24*(-T31-6*S31+U1+5*R1+B1);
*(-T32-6*S32+U2+5*R2+B2)/n/(n-1)/(n-2)/(n-3);
m3 = m3 + 3*(T11*T11*T11+2*T11*(T21-5*S21)+24*S31-8*R1-8*B1);
*(T12*T12*T12+2*T12*(T22-5*S22)+24*S32-8*R2-8*B2);
/n/(n-1)/(n-2)/(n-3)/(n-4);
m3 = m3 + 12*(T11*(T21-2*S21)+2*T31+24*S31-4*U1-16*R1-4*B1);
*(T12*(T22-2*S22)+2*T32+24*S32-4*U2-16*R2-4*B2);
/n/(n-1)/(n-2)/(n-3)/(n-4);
m3 = m3 + (-T11*T11*T11-6*T11*(T21-3*S21);
-8*T31-120*S31+16*U1+72*R1+24*B1);
*(-T12*T12*T12-6*T12*(T22-3*S22)-8*T32-120*S32+16*U2+72*R2+24*B2);
/n/(n-1)/(n-2)/(n-3)/(n-4)/(n-5);
mom[2] = m3;
var = m2-m1*m1;
variance[0] = var;
et = sqrt(var);
skew = (m3-3*m1*var-m1*m1*m1)/et/et/et;
skewness[0] = skew;
return;
}
}

## Fonction de la statistique et du test dans R
# appelant une fonction écrite en C++
dyn.load('dcovp.dll') # pour charger toutes les fonctions

```

```

# continues dans dcovp.dll
Hn2p.cpp <- function (X1, X2, index=1)
{
  if(is.vector(X1)){
    n=length(X1)
  }
  if(is.matrix(X1)){
    n=nrow(X1)
  }
  out.dcovp <- .C('dcovp', x1=t(X1), x2=t(X2), stat=as.double(0),;
    mom=rep(0,3), variance=as.double(0), skewness=as.double(0),;
    nrowx = as.integer(nrow(X1)), ncolx1=as.integer(ncol(X1)),;
    ncolx2=as.integer(ncol(X2)), index = as.double(index))
  TrA1A2 <- out.dcovp$stat*(n^2)
  mean.T <- out.dcovp$mom[1]
  variance.T <- out.dcovp$variance
  as <- out.dcovp$skewness
  TrA1A2.std <- (TrA1A2 - mean.T)/(variance.T^0.5)
  if (as >= 0)
    prob <- pgamma(TrA1A2.std - (-2/as), shape = (4/as^2),;
      scale = (as/2), lower.tail = FALSE)
  if (as < 0)
    prob = pgamma(as/abs(as) * TrA1A2.std + 2/abs(as),;
      shape = (4/as^2), scale = (abs(as)/2))
  return(list(stat = TrA1A2/(n^2), mean.T = mean.T,;
    variance.T = variance.T, skewness.T = as, p.value = prob))
}

```

Annexe B

PROGRAMME R DU DEUXIÈME ARTICLE

```
## Fonction des deux premiers moments de permutation de tr(A1...Ap),
# A est une liste de matrices A1,...Ap
perm.var.tr.A1...Ap <- function(A,p){
  n <- nrow(A[[1]])
  # Notations et fonctions de base
  G <- function(A){
    dA <- diag(A)
    T <- sum(dA)
    A2c <- A^2
    T2 <- sum(A2c)
    dA2 <- dA^2
    S2 <- sum(dA2)
    A2m <- A%*%A
    list(T=T,T2=T2,S2=S2)
  }
  T <- c()
  T2 <- c() ; S2 <- c()
  temp <- 1
  for(j in 1:p){
    temp <- temp*A[[j]]
    outG <- G(A[[j]])
    T[j] <- outG$T
    T2[j] <- outG$T2
    S2[j] <- outG$S2
  }
  # Formule du 1er moment
  mom1 <- prod(T)/(n^(p-1)) + prod(-T)/( (n*(n-1))^(p-1) )
```

```

# Formule du 2e moment
mom2 <- prod(S2)/( n^(p-1) );
  + ( prod(T^2-S2)+2*prod(T2-S2)+4*prod(-S2) );
  /( (n*(n-1))^(p-1) ) + ( 4*prod(2*S2-T2)+2*prod(2*S2-T^2) );
  /( (n*(n-1)*(n-2))^(p-1) ) + prod(2*T2-6*S2+T^2);
  /( (n*(n-1)*(n-2)*(n-3))^(p-1) )
mean.T <- mom1
variance.T <- (mom2 - mom1^2)
return(list(mean.T=mean.T,variance.T=variance.T))
}

## Fonction de la statistique et du test dans R
MultIndep.pearson <- function(X,vecd.ou.p,index=1,cte=NULL,;
method="dcov ou hsic",alpha=0.05,display=TRUE,graphics=TRUE)
{
  X <- as.matrix(X)
  n <- nrow(X)
  # Fonction de la matrice doublement centrée A
  fctA <- function(a){
    mean.ak. <- rowMeans(a)
    mean.a.l <- colMeans(a)
    mean.a <- mean(a)
    A <- sweep(a, MARGIN = 1, STATS = mean.ak., FUN = "-")
    A <- sweep(A, MARGIN = 2, STATS = mean.a.l, FUN = "-")
    A <- A + mean.a
    return(A)
  }
  # Fonction des termes pour calculer les moments de permutation
  G <- function(A){
    dA <- diag(A)
    T <- sum(dA)
    A2c <- A^2
    T2 <- sum(A2c)
    dA2 <- dA^2
    S2 <- sum(dA2)
    A2m <- A%*%A
    T3 <- sum(A2m*A)
    S3 <- sum(dA2*dA)

```

```

    U <- sum(A2c*A)
    R <- as.vector(t(dA)%*%diag(A2m))
    B <- as.vector(t(dA)%*%A%*%dA)
    list(T=T,T2=T2,S2=S2,T3=T3,S3=S3,U=U,R=R,B=B)
  }
# si length(vecd.ou.p)>1, alors cas non sériel sinon cas sériel
if (length(vecd.ou.p) > 1) {
  #- Cas non sériel
  seriel <- 0
  vecd <- vecd.ou.p
  p <- length(vecd)
  taille <- 2^p-p-1
  beta <- (1-alpha)^(1/taille)
  # Initialiser les vecteurs des outputs: stat UnA, mp1, mp2,
  # variance, mp3, skewness, pvalue
  vect.UnA <- rep(0,taille)
  mp1.UnA <- rep(0,taille)
  mp2.UnA <- rep(0,taille)
  var.UnA <- rep(0,taille)
  mp3.UnA <- rep(0,taille)
  asym.UnA <- rep(0,taille)
  pvalue.UnA <- rep(0,taille)
  vect.seuil <- rep(0,taille)
  b.norm <- rep(0,taille)
  # Extraire chaque bloc de vecteurs dans la matrice X
  X.list <- as.list(1:p)
  X.list[[1]] <- X[,1:vecd[1]]
  for(j in 2:p){
    X.list[[j]] <- X[, (sum(vecd[1:(j-1)))+1:sum(vecd[1:j])]
  }
  # Liste de toutes les matrices a1,...,ap basées sur les données X.list
  # - dcov
  a.p <- as.list(1:p)
  for(j in 1:p) a.p[[j]] <- -as.matrix(dist(X.list[[j]])^index)
  A.p <- as.list(1:p)
  # - HSIC
  if(method=="hsic"){
    fcta.b <- function(a,b){

```

```

    beta <- b/median(a[a>0])
    a <- exp( -a*(beta^index) )
    return(a)
  }
  if (is.null(cte)){
    var.perm <- function(b){
a.b <- as.list(1:p)
for(j in 1:p) {
  a.b[[j]] <- fcta.b(-a.p[[j]],b[j])
  A.p[[j]] <- fctA(a.b[[j]])
}
    -perm.var.tr.A1...Ap(A.p,p)$variance.T
  }
  b.init <- c()
  for(j in 1:p) b.init[j] <- .01/median((-a.p[[j]][-a.p[[j]]>0])^(1/index))
  out.opt <- nlm(var.perm, p=b.init)
  b.opt <- out.opt$estimate
  beta.opt <- c()
  for(j in 1:p){
    beta.opt[j] <- b.opt[j]/median((-a.p[[j]][-a.p[[j]]>0])^(1/index))
    a.p[[j]] <- fcta.b(-a.p[[j]],b.opt[j])
  }
}
}else{
  for(j in 1:p) a.p[[j]] <- fcta.b(-a.p[[j]],cte[j])
}
}

# Liste de toutes les matrices A1,...,Ap basées sur les matrices a1,...,ap
for(j in 1:p) A.p[[j]] <- fctA(a.p[[j]])
# Liste de tous les sous-ensembles A. Chaque élément de liste
# contient des sous-ensembles de même cardinal
RES <- as.list(1:p)
for(cardA in 2:p){ RES[[cardA]] <- as.matrix(combn(p,cardA)) }
# Boucle de calcul des vecteurs des outputs. Chaque élément de
# vecteur est un output pour un des sous-ensembles A, |A|>1.
nb <- 0 # Compteur pour parcourir les sous-ensembles A, |A|>1.
for(cardA in 2:p){
  for(j in 1:(choose(p,cardA))){
    nb <- nb+1

```

```

temp <- 1
T <- rep(1,p)
T2 <- rep(1,p) ; S2 <- rep(1,p)
T3 <- rep(1,p) ; S3 <- rep(1,p) ; U <- rep(1,p)
R <- rep(1,p) ; B <- rep(1,p)
for(ensA in RES[[cardA]][,j]){
  temp <- temp*A.p[[ensA]]
  outG <- G(A.p[[ensA]])
  T[ensA] <- outG$T
  T2[ensA] <- outG$T2
  S2[ensA] <- outG$S2
  T3[ensA] <- outG$T3
  S3[ensA] <- outG$S3
  U[ensA] <- outG$U
  R[ensA] <- outG$R
  B[ensA] <- outG$B
}
num.UnA <- sum(temp)
vect.UnA[nb] <- num.UnA/n # calcul de la statistique UnA
T <- T[RES[[cardA]][,j]]
T2 <- T2[RES[[cardA]][,j]] ; S2 <- S2[RES[[cardA]][,j]]
T3 <- T3[RES[[cardA]][,j]] ; S3 <- S3[RES[[cardA]][,j]]
U <- U[RES[[cardA]][,j]] ; R <- R[RES[[cardA]][,j]]
B <- B[RES[[cardA]][,j]]
# Formule du 1er moment
mp1.UnA[nb] <- prod(T)/(n^(cardA-1)) + prod(-T)/((n*(n-1))^(cardA-1))
# Formule du 2e moment
mp2.UnA[nb] <- prod(S2)/(n^(cardA-1)) ;
+ ( prod(T^2-S2)+2*prod(T2-S2)+4*prod(-S2) )/((n*(n-1))^(cardA-1)) ;
+ ( 4*prod(2*S2-T2)+2*prod(2*S2-T^2) )/((n*(n-1)*(n-2))^(cardA-1)) ;
+ prod(2*T2-6*S2+T^2)/((n*(n-1)*(n-2)*(n-3))^(cardA-1))
# Formule du 3e moment
SP1 <- prod(S3)/(n^(cardA-1))
SP2 <- ( 4*prod(-S3+U)+3*prod(T*S2-S3) + 6*prod(-S3) ;
+ 12*prod(-S3+R) + 6*prod(-S3+B) )/((n*(n-1))^(cardA-1))
SP3 <- ( 3*prod(-T*S2+2*S3) + prod(T^3-3*T*S2+2*S3) ;
+ 12*prod(-T*S2+2*S3-B) + 12*prod(2*S3-R) + 24*prod(2*S3-R-B) ;
+ 6*prod(T*(T2-S2)+2*S3-2*R) + 24*prod(2*S3-U-R) ;

```

```

+ 8*prod(T3+2*S3-3*R) )/( (n*(n-1)*(n-2))^(cardA-1) )
SP4 <- ( 12*prod(T*S2-6*S3+2*R+2*B) + 6*prod(T*(-T2+S2)-6*S3+2*U+4*R) ;
+ 3*prod(-T^3+5*T*S2-6*S3+2*B) + 12*prod(T*(-T2+2*S2)-6*S3+3*R+2*B) ;
+ 8*prod(-6*S3+2*U+3*R) + 24*prod(-T3-6*S3+U+5*R+B) );
/( (n*(n-1)*(n-2)*(n-3))^(cardA-1) )
SP5 <- ( 3*prod(T^3+2*T*(T2-5*S2)+24*S3-8*R-8*B) ;
+ 12*prod(T*(T2-2*S2)+2*T3+24*S3-4*U-16*R-4*B) );
/( (n*(n-1)*(n-2)*(n-3)*(n-4))^(cardA-1) )
SP6 <- prod(-T^3-6*T*(T2-3*S2)-8*T3-120*S3+16*U+72*R+24*B);
/( (n*(n-1)*(n-2)*(n-3)*(n-4)*(n-5))^(cardA-1) )
mp3.UnA[nb] <- SP1+SP2+SP3+SP4+SP5+SP6
var.UnA[nb] = mp2.UnA[nb] - mp1.UnA[nb]^2
cumulant3 = mp3.UnA[nb] - 3*mp1.UnA[nb]*var.UnA[nb] - mp1.UnA[nb]^3
asym.UnA[nb] = cumulant3/(var.UnA[nb]^(3/2))
UnAstd <- (num.UnA - mp1.UnA[nb])/(var.UnA[nb]^0.5)
as <- asym.UnA[nb]
if(as >= 0){
  pvalue.UnA[nb] <- pgamma(UnAstd - (-2/as), shape = (4/as^2), ;
  scale = (as/2), lower.tail = FALSE)
  vect.seuil[nb] <- ( ( qgamma(beta, shape = (4/as^2), scale = (as/2),;
  lower.tail = TRUE) + (-2/as) )*(var.UnA[nb]^0.5) + mp1.UnA[nb] )/n
}
if(as < 0){
  pvalue.UnA[nb] = pgamma(as/abs(as) * UnAstd + 2/abs(as), ;
  shape = (4/as^2), scale = (abs(as)/2))
  vect.seuil[nb] <- ( ( qgamma(beta, shape = (4/as^2), ;
  scale = (abs(as)/2), lower.tail = TRUE) ;
  + (-2/abs(as)) )*(abs(as)/as)*(var.UnA[nb]^0.5) + mp1.UnA[nb] )/n
}
names(vect.UnA)[nb] <- paste("{",paste(RES[[cardA]][,j],;
sep = ",",collapse = ","),"}")
# Afficher les outputs par sous-ensemble
if(display){ cat(c(nb,": A =",RES[[cardA]][,j],": UnA =",;
vect.UnA[nb],";
skew =",asym.UnA[nb],": seuil =",vect.seuil[nb],;
": pvalue =",pvalue.UnA[nb], "\n")) }
}
}

```



```

# Afficher le dependogram
if(graphics){
# On trace une barre verticale pour chaque A de hauteur UnA
par(mar=c(7, 4, 4, 2) + 0.1)
matplot(vect.UnA,type="h",ylim=c(0,max(c(max(vect.seuil),;
max(vect.UnA))))),;
xlim=c(0,2^p-p),main=expression(Dependogram),;
ylab="Statistic per subset",xaxt="n")
points((1:(2^p-p-1)),vect.seuil,pch="*")
axis(side=1,at=1:taille,labels=labels(vect.UnA),las = 2)
par(mar=c(5, 4, 4, 2) + 0.1)
}

# statistique et p-valeur globale
Un <- -2*sum(log(pvalue.UnA))
fisher.pvalue <- 1-pchisq(Un,2*taille)
# Les outputs
out <- list(vect.UnA=vect.UnA,seuil=vect.seuil,pvalue=pvalue.UnA,;
fisher.statistic=Un,fisher.pvalue=fisher.pvalue)
return(out)
} # fin du cas non sériel
if (length(vecd.ou.p) == 1) {
#- Cas sériel
seriel <- 1
p <- vecd.ou.p
Y <- X
d<-ncol(Y)
vecd<-rep(d,p)
m<-n-p+1
taille <- 2^(p-1)-1
beta <- (1-alpha)^(1/taille)
# Initialiser les vecteurs des outputs: stat UnA, mp1, mp2,
# variance, mp3, skewness, pvalue
vect.UnA <- rep(0,taille)
mp1.UnA <- rep(0,taille)
mp2.UnA <- rep(0,taille)
var.UnA <- rep(0,taille)
mp3.UnA <- rep(0,taille)
asym.UnA <- rep(0,taille)

```

```

pvalue.UnA <- rep(0,taille)
vect.seuil <- rep(0,taille)
# Extraire chaque bloc de vecteurs dans la matrice X
X <- matrix(0,nrow=m,ncol=p*d)
for(i in 1:m){
  for(j in 1:p){
    X[i,(((j-1)*d+1):(j*d))] <- Y[i+j-1,]
  }
}
X.list <- as.list(1:p)
X.list[[1]] <- X[,1:vecd[1]]
for(j in 2:p){
  X.list[[j]] <- X[, (sum(vecd[1:(j-1)])+1):sum(vecd[1:j])]
}
# Liste de toutes les matrices a1,...,ap basées sur les données X.list
# - dcov
a.p <- as.list(1:p)
for(j in 1:p) a.p[[j]] <- -as.matrix(dist(X.list[[j]])^index)
A.p <- as.list(1:p)
# - HSIC
if(method=="hsic"){
  fcta.b <- function(a,b){
    beta <- b/median(a[a>0])
    a <- exp( -a*(beta^index) )
    return(a)
  }
  if (is.null(cte)){
    var.perm <- function(b){
      a.b <- as.list(1:p)
      for(j in 1:p) {
a.b[[j]] <- fcta.b(-a.p[[j]],b[j])
A.p[[j]] <- fctA(a.b[[j]])
      }
      -perm.var.tr.A1...Ap(A.p,p)$variance.T
    }
    b.init <- c()
    for(j in 1:p) b.init[j] <- .01/median((-a.p[[j]][-a.p[[j]]>0])^(1/index))
    out.opt <- nlm(var.perm, p=b.init)
  }
}

```

```

b.opt <- out.opt$estimate
for(j in 1:p) a.p[[j]] <- fcta.b(-a.p[[j]],b.opt[j])
  }else{
for(j in 1:p) a.p[[j]] <- fcta.b(-a.p[[j]],cte[j])
  }
}
# Liste de toutes les matrices A1,...,Ap basées sur
# les matrices a1,...,ap
for(j in 1:p) A.p[[j]] <- fctA(a.p[[j]])
# Liste de tous les sous-ensembles A contenant 1. Chaque élément
# de liste contient des sous-ensembles de même cardinal
RES <- as.list(1:p)
for (cardA in 2:p) {
RES[[cardA]] <- as.matrix(rbind(rep(1,choose(p-1,cardA-1)),;
  as.matrix(combn(p-1,cardA-1)+1))))}
# Boucle de calcul des vecteurs des outputs. Chaque élément de
# vecteur est un output pour un des sous-ensembles A, |A|>1.
nb <- 0 # Compteur pour parcourir les sous-ensembles A, |A|>1.
for(cardA in 2:p){
  for(j in 1:(choose(p-1,cardA-1))){
    nb <- nb+1
    temp <- 1
    T <- rep(1,p)
    T2 <- rep(1,p) ; S2 <- rep(1,p)
    T3 <- rep(1,p) ; S3 <- rep(1,p) ; U <- rep(1,p)
    R <- rep(1,p) ; B <- rep(1,p)
    for(ensA in RES[[cardA]][,j]){
      temp <- temp*A.p[[ensA]]
      outG <- G(A.p[[ensA]])
      T[ensA] <- outG$T
      T2[ensA] <- outG$T2
      S2[ensA] <- outG$S2
      T3[ensA] <- outG$T3
      S3[ensA] <- outG$S3
      U[ensA] <- outG$U
      R[ensA] <- outG$R
      B[ensA] <- outG$B
    }
  }
}

```

```

num.UnA <- sum(temp)
vect.UnA[nb] <- num.UnA/m    # calcul de la statistique UnA
T <- T[RES[[cardA]][,j]]
T2 <- T2[RES[[cardA]][,j]] ; S2 <- S2[RES[[cardA]][,j]]
T3 <- T3[RES[[cardA]][,j]] ; S3 <- S3[RES[[cardA]][,j]]
U <- U[RES[[cardA]][,j]] ; R <- R[RES[[cardA]][,j]]
B <- B[RES[[cardA]][,j]]
# Formule du 1er moment
mp1.UnA[nb] <- prod(T)/(n^(cardA-1)) + prod(-T)/((n*(n-1))^(cardA-1))
# Formule du 2e moment
mp2.UnA[nb] <- prod(S2)/(n^(cardA-1)) ;
+ ( prod(T^2-S2)+2*prod(T2-S2)+4*prod(-S2) )/((n*(n-1))^(cardA-1)) ;
+ ( 4*prod(2*S2-T2)+2*prod(2*S2-T^2) )/((n*(n-1)*(n-2))^(cardA-1)) ;
+ prod(2*T2-6*S2+T^2)/((n*(n-1)*(n-2)*(n-3))^(cardA-1))
# Formule du 3e moment
SP1 <- prod(S3)/(n^(cardA-1))
SP2 <- ( 4*prod(-S3+U)+3*prod(T*S2-S3) + 6*prod(-S3) ;
+ 12*prod(-S3+R) + 6*prod(-S3+B) )/((n*(n-1))^(cardA-1))
SP3 <- ( 3*prod(-T*S2+2*S3) + prod(T^3-3*T*S2+2*S3) ;
+ 12*prod(-T*S2+2*S3-B) + 12*prod(2*S3-R) + 24*prod(2*S3-R-B) ;
+ 6*prod(T*(T2-S2)+2*S3-2*R) + 24*prod(2*S3-U-R) ;
+ 8*prod(T3+2*S3-3*R) )/((n*(n-1)*(n-2))^(cardA-1))
SP4 <- ( 12*prod(T*S2-6*S3+2*R+2*B) + 6*prod(T*(-T2+S2)-6*S3+2*U+4*R) ;
+ 3*prod(-T^3+5*T*S2-6*S3+2*B) + 12*prod(T*(-T2+2*S2)-6*S3+3*R+2*B) ;
+ 8*prod(-6*S3+2*U+3*R) + 24*prod(-T3-6*S3+U+5*R+B) ) ;
/( (n*(n-1)*(n-2)*(n-3))^(cardA-1))
SP5 <- ( 3*prod(T^3+2*T*(T2-5*S2)+24*S3-8*R-8*B) ;
+ 12*prod(T*(T2-2*S2)+2*T3+24*S3-4*U-16*R-4*B) ) ;
/( (n*(n-1)*(n-2)*(n-3)*(n-4))^(cardA-1))
SP6 <- prod(-T^3-6*T*(T2-3*S2)-8*T3-120*S3+16*U+72*R+24*B) ;
/( (n*(n-1)*(n-2)*(n-3)*(n-4)*(n-5))^(cardA-1))
mp3.UnA[nb] <- SP1+SP2+SP3+SP4+SP5+SP6
var.UnA[nb] = mp2.UnA[nb] - mp1.UnA[nb]^2
cumulant3 = mp3.UnA[nb] - 3*mp1.UnA[nb]*var.UnA[nb] - mp1.UnA[nb]^3
asym.UnA[nb] = cumulant3/(var.UnA[nb]^(3/2))
UnAstd <- (num.UnA - mp1.UnA[nb])/(var.UnA[nb]^0.5)
as <- asym.UnA[nb]
if(as >= 0){

```

```

pvalue.UnA[nb] <- pgamma(UnAstd - (-2/as), shape = (4/as^2), ;
  scale = (as/2), lower.tail = FALSE)
vect.seuil[nb] <- ( ( qgamma(beta, shape = (4/as^2), ;
  scale = (as/2), lower.tail = TRUE) ;
  + (-2/as) )*(var.UnA[nb]^0.5) + mp1.UnA[nb] )/m
}
if(as < 0){
pvalue.UnA[nb] = pgamma(as/abs(as) * UnAstd + 2/abs(as), ;
  shape = (4/as^2), scale = (abs(as)/2))
vect.seuil[nb] <- ( ( qgamma(beta, shape = (4/as^2), ;
  scale = (abs(as)/2), lower.tail = TRUE) ;
  + (-2/abs(as)) )*(abs(as)/as)*(var.UnA[nb]^0.5) + mp1.UnA[nb] )/m
}
names(vect.UnA)[nb] <- paste(" ",paste(RES[[cardA]][,j], ;
  sep = ", ",collapse = ", "),")")
# Afficher les outputs par sous-ensemble
if(display){ cat(c(nb,": A =",RES[[cardA]][,j],": UnA =",;
  vect.UnA[nb],": skew =",asym.UnA[nb],": seuil =",vect.seuil[nb],;
  ": pvalue =",pvalue.UnA[nb], "\n")) }
}
}
# Seuils critiques par sous-ensembles de même cardinal
serial.seuil <- rep(0,p-1)
begin <- 1
end <- 0
for (cardA in 2:p) {
  end <- end+choose(p-1,cardA-1)
  vecA <- vect.seuil[begin:end]
  vecA <- sort(vecA)
  serial.seuil[cardA-1] <- vecA[round(beta*choose(p-1,cardA-1))]
  begin <- end+1
}
# Afficher le dependogram
if(graphics){
# On trace une barre verticale pour chaque A de hauteur UnA
par(mar=c(7, 4, 4, 2) + 0.1)
matplot(vect.UnA,type="h",ylim=c(0,max(c(max(vect.seuil),;
  max(vect.UnA))))),;

```

```

xlim=c(0,2^(p-1)),main=expression(paste(Adaptive," ",HSIC)),;
ylab="Statistic per subset",xaxt="n")
begin <- 1
end <- 0
for (cardA in 2:p) {
  end <- end+choose(p-1,cardA-1)
  segments(begin-0.5,serial.seuil[cardA-1],end+0.5,;
  serial.seuil[cardA-1],lty=4)
  begin <- end+1
}
axis(1,1:taille,labels(vect.UnA),las=2)
par(mar=c(5, 4, 4, 2) + 0.1)
}
# statistique et p-valeur globale
Un <- -2*sum(log(pvalue.UnA))
fisher.pvalue <- 1-pchisq(Un,2*taille)
# Les outputs
out <- list(RES=RES,vect.UnA=vect.UnA,seuil=vect.seuil,;
pvalue=pvalue.UnA,fisher.statistic=Un,fisher.pvalue=fisher.pvalue)
return(out)
} # fin du cas sériel
# fin du programme
}

```